THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Changing the Way We Measure Time to More Accurately Estimate the Probability of Informed Trading

Xin Ling

Master of Financial Management

Master of Professional Accounting

Bachelor of Science

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2016

UQ Business School

Abstract

For decades, finance researchers have been interested in the distribution of stock returns. The empirical evidence of prior studies does not support a normal distribution of stock returns; however, we manage to observe a normal distribution of stock returns on an intraday basis by changing the way to measure time with an event clock. A normal distribution of stock returns is a common underlying assumption for developing financial models in areas such as asset pricing and market microstructure. Our findings on the distribution of returns based on an event clock fulfil the assumption of normality. Therefore, there is benefit in applying our method in such studies and findings based on our event clock should be more accurate and reliable. Motivated by the impact of significant change in market conditions on the behaviour of market makers, we apply the method to examine a bid-ask spread model proposed by Bollen, Smith and Whaley (2004) (BSW), which assumes the normality of stock returns. Our work sheds light on the bid-ask spread cost components of market makers in the current high-frequency market; we also find that the explanatory power of the model is significantly improved by applying the event clock setting. Based on the improved model, we develop new methods to estimate the probability of informed trading (PI) on a market level and a stock level. Our methods of PI estimation allow us to identify PI of a single stock, as well as the market, on a daily basis providing powerful tools for investors, regulators and researchers to monitor informed trading around significant events.

The thesis contains three essays that address the issues described above. The first essay introduces a new way to measure time using event clocks, which is different from the "default" time measurement of most finance studies, which is the calendar clock. We show that our event clocks outperform the calendar clock in capturing the level of market activity. We then examine the intraday stock returns distribution using a calendar clock versus event clocks. We find that returns do not follow a normal distribution with a traditional calendar clock, but do follow a normal distribution when event clocks, especially the transaction clock, are applied.

The findings of the first essay suggest that our transaction clock is able to uncover a normal distribution of stock returns. Based on that, we expect there to be benefit in applying the transaction clock in studies that assume a normal distribution of stock returns.

O'Hara (2015) challenges the validity of classic market microstructure models in the current high-frequency market; nevertheless, in the second essay, we apply our transaction clock method to re-evaluate the determinants of a market maker's bid-ask spread under current market conditions, by examining the BSW model, which assumes the normality of stock returns. We attempt to answer two questions: 1) What is the impact of the high-frequency market on the bid-ask spread cost components; and 2) How does the application of a transaction clock improve a model with a normality assumption. We conduct empirical tests with a traditional calendar clock versus a transaction clock. We find that inventory costs and adverse selection costs still have significant impact on the bid-ask spread, while order processing costs are not as significant. This result can be attributed to the impact of the current high-frequency market. The BSW model possesses good explanatory power for the market makers' bid-ask spreads in the current high-frequency market; however, we find that applying a transaction clock significantly improves the model's explanatory power. Based on the improved model, we develop a method to estimate probability of informed trading (PI) on a market level using a restricted regression with intraday observations. Our proposed method allows us to identify informed trading over short periods, such as an hour, a day and a week.

The third essay examines the impact of the global financial crisis (GFC) on market makers' behaviours and on adverse selection in stock trading. At a moment of crisis, an uninformed trader will leave the market due to concern that other uninformed traders will exit and he/she will be left trading with informed traders. With the exit of uninformed traders, the market will be dominated by informed traders, which will result in increased adverse selection during the crisis period. We modify the BSW model with period dummies to test the impact of the GFC on bid-ask spread cost components and to examine the PI on a market level during the crisis and non-crisis periods using the method from the second essay. Moreover, we develop a new method to estimate daily PI of individual stocks, which allows us to estimate the PI on a stock level in different periods. Our findings suggest significantly higher adverse selection on both the market level and stock level during the crisis period compared to the non-crisis period. We also find market makers tend to be more conservative in setting their bid-ask spreads during the crisis period.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Ling, X., 2015. Normality of stock returns with event time clocks. *Accounting & Finance*. Accepted 15 March 2015.

Chen, X., and Ling, X., 2015. Determinants of Chinese equity financing behaviours: traditional model and the alternatives. *Accounting & Finance*. Accepted 15 March 2015.

Publications included in this thesis

Ling, X., 2015. Normality of stock returns with event time clocks. *Accounting & Finance*. - incorporated as Chapter 2.

Contributor	Statement of contribution
Ling, X. (Candidate)	Designed experiments (100%) Wrote the paper (100%)

Contributions by others to the thesis

No contributions by others.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

I would like to take this opportunity to express my gratitude to those who have been supporting me since I started my PhD candidature.

First, I am deeply indebted to my supervisor, Professor Tom Smith, who has been providing me with all kinds of support, inspiration, encouragement and unconditional love for the past several years. Without his insightful advice and enormous help, this thesis would never have been started and completed. I have been extremely fortunate to have a supervisor who cares about me as well as my work, and who always responds promptly whenever I need his help. Tom is like a second father to me. I would never have reached where I am today without his guidance.

I would like to thank my associate advisors, Dr. Elizabeth Zhu and Dr. Martina Linnenluecke for their support and encouragement throughout the years of my PhD. They've always been nice to me and provided me with helpful advice. I also want to thank Professor Robert Faff and Professor Allan Hodgson for being my readers and providing valuable comments and suggestions. I would also thank Professor Karen Benson for being the Chair of the Examination Committee.

I would also like to thank the UQ Business School for providing me with scholarships, research funding and a great working environment. I've had a good time during my PhD candidature. Being surrounded by so many friendly staff and lovely friends makes this journey full of joy.

Finally, I must express my gratitude to my family. I would like to thank my parents, who have been unconditionally loving and supporting me for the past 29 years. I would also like to thank my lovely wife, Caroline, who has kept me company and helped me make it through all the tough times in the past several years.

Keywords

normality, event clock, intraday returns distribution, bid-ask spread, inventory holding premium, probability of informed trading, adverse selection, global financial crisis

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 150201, Finance, 100%

Fields of Research (FoR) Classification

FoR code: 1502, Banking, Finance and Investment, 100%

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATM	At-the-money
BSW	Bollen, Smith and Whaley
EWQS	Equal-weighted quoted spread
GFC	Global Financial Crisis (07/2007-06/2009)
IHP	Inventory holding premium
IHP_I	Inventory holding premium of a transaction with an informed trader
IHP_U	Inventory holding premium of a transaction with an uninformed trader
ITM	In-the-money
PI	Probability of informed trading
OTM	Out-of-the-money
VWES	Volume-weighted effective spread

Chapter 1

INTRODUCTION

1.1 Overview

In this thesis, we first change the way to measure time using event clocks instead of a traditional calendar clock and examine the distribution of intraday stock returns with event clocks verses the traditional calendar clock. We find that the transaction clock, in which the event is defined as the occurrence of a constant number of transactions, possesses good characteristics for capturing a constant level of market trading activity and uncovering a normal distribution of intraday stock returns. Several important financial models commonly assume the normality of stock returns distribution; therefore, there is benefit in applying the transaction clock in finance studies, as it fulfils the underlying assumption of normality.

We then re-evaluate the bid-ask spreads cost components in the current high-frequency market on an intraday basis, motivated by the impact of the significant change in market conditions. To conduct empirical tests, we employ Bollen, Smith and Whaley's (2004) (BSW) bid-ask spread model, which assumes a normal distribution of stock returns. We perform the tests with a transaction clock versus a calendar clock. The transaction clock is found to outperform the calendar clock in improving the explanatory power of the model. In addition, consistent evidence suggests the BSW model possesses good explanatory power, despite the significant change in market conditions.

Based on the BSW model, we develop two new methods to estimate the probability of informed trading (PI) on a market level as well as a stock level. We apply our methods of PI estimation to examine the impact of the recent global financial crisis (GFC) on the behaviour of market makers and on adverse selection in stock trading. We find that market makers are more conservative in the GFC period, and observe a significant increase in adverse selection in the GFC period on a market level and a stock level.

The remainder of this thesis is structured as follows: Chapter 2 introduces the way to measure time using event clocks and examines the distribution of intraday stock returns with event clocks versus a calendar clock. Chapter 3 applies the transaction clock to re-evaluate the bid-ask spread components in the high-frequency market on an intraday basis, and develops a method to estimate PI on a market level. Chapter 4 develops a method to estimate PI on a stock level, and applies this method, together with the method from Chapter 3, to examine the impact of the GFC on market makers' behaviours and on adverse selection in the stock trading. Chapter 5 concludes the thesis.

1.2 Distribution of Stock Returns

For decades, researchers have explored the behaviour of stock prices in association with market activity and much effort has been devoted to determining the distribution of returns. The cumulative evidence suggests that daily returns of various types of assets do not follow a normal distribution (Mandelbrot 1963; Fama 1965; Upton and Shannon 1979; Affleck-Graves and McDonald 1989). The current available high-frequency trading data allows several tests of stock returns distribution to be performed on an intraday basis, and the findings do fully not support a normal distribution of stock returns (Andersen et al. 2001; Ané and Geman 2000; Easley et al. 2012).

Despite empirical evidence against the normal distribution of stock returns, questions have been raised about the choice of a proper clock to measure time in finance studies. Clark (1973) suggests that the correct measure of time change in financial markets should represent a constant level of trading activity rather than a constant length of time. Similarly, Mandelbrot and Taylor (1967) suggest that price changes over a fixed number of transactions may follow a normal distribution. On the other hand, the random walk hypothesis suggests a stock price should follow a random walk, implying a log-normal distribution of price changes. This is rejected by Lo and Mackinlay (1988), within the calendar clock setting. Since the hypothesis itself does not specify a measure of time change, the rejection of the hypothesis can result from the choice of a calendar clock. Therefore, it becomes a matter of time change measurement, with which the random walk theory stands, as well as the normal distribution of stock returns. In this thesis, we address the issue of choosing a proper clock, and introduce a different way to measure time using event clocks.

1.3 Characteristics of Event Clocks

Although prior studies do not support a normal distribution of stock returns, they reveal that the choice of an appropriate clock can be critical to determining the distribution of stock returns. With a proper event clock, we expect stock returns to follow a distribution that is closer to a normal distribution than with a calendar clock. In this thesis, we extend the work of Clark (1973), Ané and Geman (2000) and Easley et al. (2012) and use event clocks as time change measurements to obtain a constant level of market activity, rather than time length. We define the first type of events to be the occurrence of a certain number of transactions and the second type of events to be the occurrence of a certain amount of trading volume. With the available high-frequency trading data, our event clocks, termed the “transaction clock” and the “volume clock”, can be clearly identified on an intraday basis. Since transactions and trading volume capture the level of trading activity, the event clocks built on these two types of events fulfil the condition that the time changes over event clocks contain a constant level of trading activity. We expect to observe a close-to-normal distribution of intraday stock returns within the setting of event clocks.

Chapter 2 introduces the way to use a transaction clock and a volume clock to measure time and conducts theoretical and empirical analysis on the distribution of intraday stock returns. We first model stock price changes with event clocks using a Brownian motion. The model implies a normal distribution of returns with event clocks and a non-normal distribution with a calendar clock. We then empirically test the intraday returns distribution of Dow Jones 30 (DJ30) stocks using the Kernel Method and Hansen’s (1982) generalized method of moments (GMM). Despite mixed findings with a calendar clock and a volume clock, we consistently observe a normal distribution of intraday stock returns with a transaction clock. Our findings suggest that the transaction clock possesses good characteristics, such as capturing a constant level of market activity and uncovering a normal distribution of intraday stock returns. There is benefit in applying the transaction clock in finance studies, especially those that assume a normal distribution of stock returns. We apply our transaction clock in Chapter 3 as an example, and illustrate that the transaction clock outperforms the calendar clock in achieving higher explanatory power of the model.

1.4 Bid-Ask Spread Determinants

Motivated by the impact of significant change in market conditions, Chapter 3 re-evaluates the bid-ask spread components in the current high-frequency market on an intraday basis.

The bid-ask spread of a market maker is regarded as cost of immediacy (Demsetz 1968). Analysing the bid-ask spread and its cost components allows us to understand market designs and market makers' behaviours. Stoll (1978) developed a theoretical model for the bid-ask spread with three major cost components; namely, order processing costs, inventory holding costs and adverse selection costs. Since then, various studies have established a detailed understanding of the cost components of the bid-ask spread.

However, market conditions have changed dramatically since that time. During the last decade, we have witnessed significant increases in daily transactions, trading volume, as well as in the number of bid and ask quotations placed by market makers. O'Hara (2015) illustrates the impact of the current high-frequency market on market microstructure theories, and challenges the validity of classic bid-ask spread models under the current market conditions. With this in mind, it is important to re-evaluate the cost components of the bid-ask spread under current market conditions, which allows us to understand the impact of the high-frequency market on the bid-ask spreads of market makers.

We employ Bollen, Smith and Whaley's (2004) (BSW) bid-ask spread model as our base model to examine the bid-ask spread cost components in the current high-frequency market on an intraday basis. The BSW model uses an at-the-money (ATM) call option to model inventory holding premium (IHP) as a proxy of inventory holding costs and adverse selection costs. At the time, the model had high explanatory power on the spreads of a cross-section of stocks. We employ this model in our study for several reasons. First, compared to most other proxies of bid-ask spread cost components, the variables constructed in the BSW model are less sensitive to significant increases in market activity. Second, the underlying assumption in studying the behaviour of market makers to minimise inventory risk is not likely to be violated by changing market conditions. Third, while the model assumes a normal distribution of stock return, applying a transaction clock would fulfil the underlying assumption and provide more accurate and reliable results.

We conduct empirical tests on frequently traded stocks in NYSE and NASDAQ within the settings of intraday intervals using the calendar clock versus the transaction clock. We find the BSW model possesses good explanatory power for the intraday bid-ask spread components in a high-frequency market. Inventory costs and adverse selection costs have significant impact on the bid-ask spread of market makers; however, order processing costs become insignificant in the current high-frequency market. Moreover, applying a transaction clock is beneficial, because it improves the model's explanatory power, compared to a calendar clock.

1.5 Probability of Informed Trading

Finance researchers have paid a lot of attention to information asymmetry in the stock market. Several rational expectation models have been constructed to account for the effect of private information (Grossman 1976; Admati 1985; Kyle 1985; Admati and Pfleiderer 1988; Foster and Viswanathan 1990; Boulatov and George 2013; and Rosu 2015). As private information leads to informed trading activity, the term "probability of informed trading" has been introduced to assess the level of information asymmetry. From a market maker's perspective, delivering a transaction initiated by an informed trader will cause a loss, which is regarded as an adverse selection cost. A market maker compensates such cost from his bid-ask spread; so that his bid-ask spread reflects his understanding of the level of information asymmetry in the transactions. Therefore, the probability of informed trading can be directly inferred from the bid-ask spreads of market makers (Easley and O'Hara 1987; BSW 2004; and Easley et al., 2012).

The BSW (2004) introduces the IHP to measure adverse selection costs and inventory holding costs. Given the chance to trade with informed traders, a part of the IHP can be attributed to transactions with informed traders, and the rest can be attributed to transactions with uninformed traders. Therefore, by decomposing the IHP into the IHPs of transactions with informed traders and uninformed traders, the expected IHP can be treated as a probability weighted average of the informed and uninformed components. The probability of informed trading can be inferred from the relation between the IHP and its components.

Based on the modified BSW model, we develop two new methods to estimate PI on a market level as well as a stock level. In Chapter 3, we use a restricted regression to estimate PI on a market level with intraday observations. In Chapter 4, we construct a fully identified equation, and solve for daily PI of individual stocks. Our proposed methods provide significant new tools to regulators, researchers, investors and other parties to understand market behaviour and deal with informed trading. As an example, in Chapter 4 we apply our methods of PI estimation to examine the impact of the GFC on the behaviours of market makers and on adverse selection in stock trading. We illustrate that our PI estimation performs well in the study and provides sensitive results.

1.6 Adverse Selection in the Global Financial Crisis

In Chapter 4, we examine the impact of a financial crisis on the behaviours of market makers and on adverse selection in stock trading. We are motivated by the severe and worldwide market crashes that occurred during the GFC, and we attempt to explain market overreaction in a financial crisis.

Morris and Shin (2012) argue that stock market crashes can be attributed to investors losing confidence and worrying about being adversely selected by informed traders in a financial crisis. Uninformed traders participate in the market when they believe trades are mutually beneficial, despite the loss when they are trading with informed traders. When a shock arrives, an uninformed trader may be concerned that other uninformed traders will exit and leave him/her trading with informed traders; thus, he/she will be less motivated to trade in the market. With the exit of uninformed traders, market participants will be dominated by informed traders, which will result in an increase in adverse selection during the crisis period.

We first employ the BSW model as a base model and modify it with period dummies to examine the impact of the GFC on market makers' behaviours. Our findings suggest that market makers tend to be more conservative during the crisis period, as they significantly widen their bid-ask spread and demand more compensation for the costs they are bearing. Moreover, market makers demand significant compensation for their inventory costs and adverse selection costs, but little for order processing costs.

We then apply the methods developed in the thesis to estimate PI on a market level and a stock level during the sample period. As is well documented in market microstructure literature, PI is a good indicator of information asymmetry and adverse selection.

Estimating PI during crisis and non-crisis periods allows us to examine the impact of the GFC on adverse selection in stock trading. We find consistent evidence for a significant increase in adverse selection during the crisis period on a market level as well as a stock level. Our findings support Morris and Shin's (2012) argument and provide an explanation for market overreaction in a financial crisis.

Chapter 2

NORMALITY OF STOCK RETURNS WITH EVENT TIME CLOCKS

2.1 Introduction

Over the past few decades, research has been conducted to explore the behaviour of stock prices associated with market activity. On one hand, researchers have been interested in the distribution of stock price change in terms of the size of the change, and its volatility and higher moments, such as skewness, kurtosis, etc. Other scholars have investigated the relations among price changes and other market factors, such as trading volume, liquidity, information, etc.

Much effort has been devoted to determining the distribution of returns. The cumulative evidence suggests that the daily returns of various asset types do not follow a normal distribution. Mandelbrot (1963) finds that the distribution of stock price changes does not follow a normal distribution, but follows a stable Paretian distribution, which has longer tails than a normal distribution. Fama (1965) finds that daily stock returns have a distribution more long-tailed than a normal distribution, but monthly returns have a distribution closer to normal. Upton and Shannon (1979) test stock return distributions over monthly, quarterly and annual periods, and find returns follow a distribution closer to normal over longer periods than over shorter periods. Affleck-Graves and McDonald (1989) have observed non-normality in monthly returns for individual stocks. Similar evidence against normality has also been found in equity markets outside the U.S. (Praetz and Wilson 1978, Stokie 1982, Brailsford 1996, Gray et al. 1998, and Doan et al. 2014).

Despite the empirical evidence that stock price changes do not follow a normal distribution, Mandelbrot and Taylor (1967) suggest that price changes over a fixed number of transactions may follow a Gaussian distribution and price changes over a fixed time period may follow a stable Parentian distribution. Similarly, Clark (1973) states that the calendar clock may not be the appropriate measure of time change in financial markets, and suggests that a normal distribution can be observed with a proper measure of time change representing a constant level of trading activity. Moreover, the random walk hypothesis suggests a stock price should follow a random walk, which would result in a log-normal distribution of price changes. Lo and Mackinlay (1988) rejected the random walk hypothesis, based on the empirical evidence in the calendar clock setting. However, the random walk theory doesn't specify a measurement of time change. It becomes a matter of time change measurement, with which the random walk theory stands, as well as the normality of stock price changes.

Our objective is to introduce a different way to measure time, with which a normal distribution of stock price changes can be observed. The proposed time change measurement employs the event clock, based on an idea from Clark (1973). The events we choose, such as the occurrence of a certain number of transactions and the occurrence of certain amount of trading volume, better represent a constant level of trading activity than the traditional calendar clock. We apply the Kernel method and Hansen's (1982) generalized method of moments (GMM) to examine the intraday stock returns distribution using different clocks. Based on the empirical evidence, we observe a close-to-normal distribution in the context of event clocks.

2.1.1 Stock Price Modelling

Over a century ago, Bachelier (1900) introduced Brownian motion for stock price modelling. It is suggested that in stochastic time, stock price changes follow the following form:

$$\mathrm{dS} = \mu(\mathrm{S})\mathrm{dt} + \sigma(\mathrm{S})\mathrm{dW_t} \tag{2.1}$$

where $\mu(S)$ is the drift term on stock price, $\sigma(S)$ is the diffusion term on stock price and W_t is a standard Brownian motion.

This stock price modelling has been adopted in several asset pricing models. Black and Scholes (1973) use the same setting for stock prices to derive the option pricing model. Merton (1973), Breeden (1979) and Cox, et al. (1985) have applied similar settings in their model derivations.

Lo and Mackinlay's (1988) test for random walk hypothesis obtains a series of discrete-time stock price changes by sampling the above continuous-time process at equally spaced intervals. The model in discrete-time has the form:

$$\Delta S = \mu(S)\Delta t + \sigma(S)\Delta W \tag{2.2}$$

As is described in the model, the equally spaced intervals follow the traditional calendar clock, and have a constant length of time Δt. However, stock price changes are more likely to be driven by trading activity than time itself. Price is unlikely to change in a particular period of Δt when there are no transactions, but is likely to be volatile in a period of Δt when there are a lot of transactions. Our proposed measurement of time change employs the event clock, instead of the calendar clock. The intervals therefore will be equally spaced with events rather than time. By doing so, we will be able to model stock price changes within the setting of event clocks.

2.1.2 Intraday Return Distribution

Several studies have explored intraday returns distribution, using available high-frequency trading data. Andersen et al. (2001) find that daily returns of 30 Dow Jones Industrial Average (DJIA) stocks follow a distribution with fatter tails than a normal distribution. However, the standardized intraday returns over 5-minute intervals follow a distribution close to normal. Ané and Geman (2000) gather the intraday returns of two particular stocks over narrow intraday calendar clock intervals over 12 months and apply classical Kernel methods to estimate the empirical density of intraday stock returns over 1-minute intervals, 5-minute intervals, 10-minute intervals and 15-minute intervals. The empirical distributions contain fatter tails and higher kurtosis compared with a normal distribution. However, the estimated density of the same series of returns conditional on the re-centred number of transactions shows a nearly normal distribution. Ané and Geman (2000) claim that the normality of stock returns can be recovered by using the number of trades as a measurement of time change, which is consistent with Clark's (1973) statement. Easley et al. (2012) recently introduced another way to measure time changes using a volume clock to identify information flow and estimate the probability of informed trading on a daily basis. Several studies also focus on intraday stock returns in Asia Pacific equity markets; their findings do not support a normal distribution of intraday stock returns within the setting of a traditional calendar clock (Cheung et al. 1994, Choe and Shin 1993, Hodgson 1996 and Ke et al. 2004).

Previous studies do not support a perfectly normal distribution of stock returns; however, they reveal that the choice of an appropriate clock is critical to determining the distribution of stock returns. With a proper event clock, we expect intraday stock returns to follow a distribution that is much closer to a normal distribution than with a calendar clock. We extend the work of Clark (1973), Ané and Geman (2000) and Easley et al. (2012) and conduct our empirical tests on intraday stock returns distribution in a more direct way. Instead of adjusting the factors of trading activity in fixed time periods, we use event clocks as time change measurements to fix the level of market activity, rather than time length. We define our first type of event to be the occurrence of a certain number of transactions and our second type of event to be the occurrence of a certain amount of trading volume. The event clocks built on these two types of events are expected to fulfil the condition that the time changes over an event clock contain a constant level of trading activity. We expect to observe a close-to-normal distribution of intraday stock returns within the setting of event clocks.

The remainder of this chapter is structured as follows: Section 2.2 introduces the methods we use to construct intraday intervals and examine intraday return distribution; Section 2.3 describes the data; Section 2.4 presents and discusses the empirical results; and Section 2.5 concludes the Chapter.

2.2 Methodology

2.2.1 Intraday Intervals with Calendar Clock and Event Clock

Calendar clock measures time changes with time length; this is regarded as a "default" measurement for time changes in finance research. Within the settings of a calendar clock, intervals are equally spaced with a constant time length. A common use of calendar clock is to compute the rate of return for a security over a day, a week, a month, a quarter and a year. In our study, we form two calendar clock intervals: five minutes and ten minutes. The first interval of a day contains the first five minutes (or ten minutes) after the market opens, and each interval contains five minutes (or ten minutes) until the market closes. Since there are 6.5 trading hours in a trading day, the number of intervals within a trading day will constantly be 78 with 5-minute intervals, and 39 with 10-minute intervals.

Event clock provides a different way to measure time changes, based on the occurrence of certain events. Following the conjecture in Clark (1973), we choose events that better represent a constant level of trading activity.

Our first type of event is defined as the occurrence of a certain number of transactions. Instead of using the same time length to determine intraday intervals, we use the same number of transactions. We form the transaction clock intervals to contain the same number of transactions (e.g. 100-transaction interval, 200-transaction interval, etc.). Instead of being equally spaced with a constant time length, transaction clock intervals are equally spaced with a constant number of transactions. The first interval of a day contains the first 100 transactions (or 200 transactions) after the market opens, and each interval contains the same number of transactions. Unlike a calendar clock, the number of intervals within a trading day varies because the number of transactions is unlikely to be the same across different stocks and trading days. Ané and Geman (2000) examine the distribution of intraday stock returns by estimating density of calendar clock stock returns conditional on the re-centered number of transactions. Our approach provides a direct way to examine the distribution of intraday stock returns using a transaction clock.

Our second type of event is defined as the occurrence of a certain amount of trading volume. We follow Easley et al.'s (2012) approach to use the same amount of trading volume to determine intraday intervals. Similar to transaction intervals, the volume clock intervals contain the same amount of trading volume (e.g., 20,000-volume interval, 40,000-volume interval, etc.). For the same reason, a volume clock would not provide a constant number of intervals within a trading day. In the case that a single transaction occurs to purchase or sell a number of shares exceeding the interval size, we divide the transaction into a number of "sub-transactions" to fit within volume intervals. The transaction volume will be the same as the cumulated volume in the allocated intervals.

2.2.2 Modeling Intraday Stock Returns with Event Clocks

We start with a stochastic stock price change with a simple Brownian motion. The price of an underlying asset, typically a stock, follows a geometric Brownian motion. In an infinitesimal period of time, the infinitesimal rate of return on the stock has an expected return of μdt and a variance of $\sigma^2 dt$. That is

$$\frac{dS}{S} = \mu dt + \sigma dW_t \tag{2.3}$$

where W is a stochastic variable (Brownian motion).

Within the setting of calendar clock intervals, each intraday interval has the same time length 'Δt'. The stock returns over each interval '$\frac{\Delta S}{S}$' with a constant time change 'Δt' is modeled as:

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma \Delta W_t \tag{2.4}$$

Within the setting of event clock intervals, the intraday event clock intervals no longer have the same time change Δt (see section 2.1.1). Instead, the i-th event clock interval contains a time change Δt_i $(i = 1,2,3, \ldots N)$, and Δt_i does not necessarily equal Δt_j when $i \neq j$.

We use 100-transaction intervals as an illustration. Δt_1 denotes the time taken for the first 100 transactions (1st interval) to occur in the market after it opens, while Δt_2 denotes the time for the following 100 transactions to occur. Within such settings, the stock returns over the i-th interval '$\frac{\Delta S_i}{S}$' can be modelled as:

$$\frac{\Delta S_i}{S} = \mu \Delta t_i + \sigma \Delta W_{t_i} \tag{2.5}$$

where ΔS_i is the change in stock price in the i-th interval.

Since Δt_i is neither constant nor certain, we aim to rewrite the model by replacing the inconstant Δt_i with the constant terms defined in the event clock; i.e. number of transactions in the transaction clock and amount of trading volume in the volume clock. To achieve this, we require an additional assumption. It is reasonable to assume that price changes are triggered by transactions or trading volume because price changes are driven by market activity, which are better represented by transactions and trading volume.

Within the setting of a transaction clock, in which each interval involves the same number of transactions, we assume that price changes are triggered by transactions. We can replace t (time) with N (number of transactions) to get a new Brownian motion:

$$dX = \mu(X)dN + \sigma(X)dW_N \tag{2.6}$$

where dX is the size of price change.

By accumulating transactions from n to m, we will have:

$$X(m) = X(n) + \int_n^m \mu(X)\, dN + \int_n^m \sigma(X)\, dW_N \tag{2.7}$$

Under Ito's lemma, with $g(X, n) = \ln X$,

$$\begin{aligned} dg &= \left[g_t + g_X \mu(X) - \frac{1}{2} g_{XX} \sigma^2(X)\right] dN + g_X \sigma(X) dW_N \\ &= \left[\frac{1}{X}\mu(X) - \frac{1}{2X^2}\sigma^2(X)\right] dN + \frac{1}{X}\sigma(X) dW_N \end{aligned} \tag{2.8}$$

Again, accumulating transactions from n to m, we will have:

$$\begin{aligned} g(X, m) &= g(X, n) + \int_n^m \frac{1}{X}[\mu(X) - \frac{1}{2X^2}\sigma^2(X)] dN + \int_n^m \frac{1}{X}\sigma(X) dW_N \\ &= g(X, n) + \frac{1}{X}\mu(X)(m - n) - \frac{1}{2X^2}\sigma^2(X)(m - n) + \frac{1}{X}\sigma(X)(W_m - W_n) \end{aligned} \tag{2.9}$$

Let $\mu(X) = \alpha X$, $\sigma(X) = \beta X$ and move $g(X, n)$ to the left side, we get:

$$\ln \frac{X(m)}{X(n)} = \left(\alpha - \frac{1}{2}\beta^2\right)(m - n) + \beta(W_m - W_n) \tag{2.10}$$

The left side of the equation stands for the log return of a stock from transaction n to transaction m. To fit in our interval setting, we choose transaction n to be the first transaction of the interval and transaction m to be the last transaction of the interval. Since each transaction clock interval contains the same number of transactions, we have $m - n$ as a constant, which is exactly the size of the interval. From the properties of Brownian motion, the return over the period between transaction n and transaction m follows a normal distribution $N[\left(\alpha - \frac{1}{2}\beta^2\right)(m - n), \beta^2(m - n)]$. Therefore, under the assumption that the price changes are triggered by transactions, the stock returns over transaction clock intervals should be normally distributed.

Within the setting of a volume clock, in which each interval involves the same amount of trading volume, we assume that price changes are triggered by trading volume. We can replace t (time) with V (trading volume) to get a new Brownian motion:

$$dX = \mu(X) dV + \sigma(X) dW_V \tag{2.11}$$

where dX is the size of price change.

Similarly, by accumulating trading volume from v to u, we will have:

$$X(u) = X(v) + \int_{v}^{u} \mu(X)\, dV + \int_{v}^{u} \sigma(X)\, dW_V \tag{2.12}$$

Under Ito's lemma, with $g(X, v) = \ln X$,

$$\begin{aligned} dg &= \left[g_t + g_X \mu(X) - \frac{1}{2} g_{XX} \sigma^2(X)\right] dV + g_X \sigma(X) dW_V \\ &= \left[\frac{1}{X}\mu(X) - \frac{1}{2X^2}\sigma^2(X)\right] dV + \frac{1}{X}\sigma(X) dW_V \end{aligned} \tag{2.13}$$

Again, accumulating trading volume from v to u, we will have:

$$\begin{aligned} g(X, u) &= g(X, v) + \int_{v}^{u} [\frac{1}{X}\mu(X) - \frac{1}{2X^2}\sigma^2(X)] dV + \int_{v}^{u} \frac{1}{X}\sigma(X) dW_V \\ &= g(X, v) + \frac{1}{X}\mu(X)(u - v) - \frac{1}{2X^2}\sigma^2(X)(u - v) + \frac{1}{X}\sigma(X)(W_u - W_v) \end{aligned} \tag{2.14}$$

Let $\mu(X) = \alpha X$, $\sigma(X) = \beta X$ and move $g(X, v)$ to the left side, we get:

$$\ln \frac{X(u)}{X(v)} = \left(\alpha - \frac{1}{2}\beta^2\right)(u - v) + \beta(\, W_u - W_v) \tag{2.15}$$

Similarly, the left side of the equation stands for the log return of a stock from trading volume v to trading volume u. To fit in our interval setting, we choose trading volume v to be the first share traded in the interval and trading volume u to be the last share traded in the interval. Since each volume clock interval contains the same number of shares traded, we again have $u - v$ as a constant, which is also the size of the interval. From the properties of Brownian motion, the stock return over this period follows a normal distribution $N[\left(\alpha - \frac{1}{2}\beta^2\right)(u - v), \beta^2(u - v)]$. Therefore, under the assumption that the price changes are triggered by trading volume, the stock return over volume clock intervals should be normally distributed.

Applying the same approach on calendar time intervals, between time t_{i-1} to t_i

$$\ln \frac{X(t_i)}{X(t_{i-1})} = \left(\alpha - \frac{1}{2}\beta^2\right) n_i + \beta(\, W_{t_i} - W_{t_{i-1}}) \tag{2.16}$$

where n_i is the number of transactions occurred between t_{i-1} and t_i;

Or

$$\ln\frac{X(t_i)}{X(t_{i-1})} = \left(\alpha - \frac{1}{2}\beta^2\right) v_i + \beta(W_{t_i} - W_{t_{i-1}}) \tag{2.17}$$

where v_i is the trading volume between t_{i-1} and t_i.

From the properties of Brownian motion, the term $\beta(W_{t_i} - W_{t_{i-1}})$ follows a normal distribution $N(0, \beta\Delta t)$, where Δt is the time length between time t_{i-1} and t_i. However, the number of transactions and the amount of trading volume is likely to be different in each time interval. Thus, $\left(\alpha - \frac{1}{2}\beta^2\right) n_i$ actually follows the distribution of n_i, while $\left(\alpha - \frac{1}{2}\beta^2\right) v_i$ actually follows the distribution of v_i. Eventually, with calendar time intervals, the stock return follows a mixture of a normal distribution and a distribution of market activity (e.g. transactions (n_i) and trading volume (v_i).

2.2.3 Kernel Analysis

To observe the empirical distribution of our intraday return sample, we apply the Kernel method to estimate the density of the empirical distribution. The Kernel estimator is:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{2.18}$$

where n is the number of observations, x_i is the i-th observation, h is the window width, and $K(\cdot)$ is a Kernel function, which is usually a symmetric probability density function. The Kernel function $K(\cdot)$ and the window width h are chosen according to a standard approach in Silverman (1986). We choose the Kernel function:

$$K(x) = (1/\sqrt{2\pi})e^{-x^2/2} \tag{2.19}$$

so that $\hat{f}$ would be a smooth curve with derivatives of all orders. We choose the ideal window width:

$$h = (\frac{4\hat{\sigma}^5}{3n})^{1/5} \approx 1.06\hat{\sigma}n^{-1/5} \tag{2.20}$$

where $\hat{\sigma}$ is the standard deviation of the sample. The estimated Kernel density would provide a visual view of the empirical distribution of the intraday return samples.

2.2.4 GMM Test of Normality

We applied Hansen's (1982) generalized method of moments (GMM) to estimate the first four moments of the intraday return sample and to test the normality of the intraday return distribution.

We compute the intraday interval returns (r_i) within the settings of calendar clock, transaction clock and volume clock. With each series of intraday return, we form the moment conditions with the first four moments of the intraday return sample.

$$g_T(\mu, \sigma^2) = \frac{1}{T}\sum_{i=1}^{T}\begin{pmatrix} r_i - \mu \\ (r_i - \mu)^2 - \sigma^2 \\ (r_i - \mu)^3 - m_3 \\ (r_i - \mu)^4 - m_4 \end{pmatrix} \tag{2.21}$$

The moment conditions are fully-identified, and we are able to estimate the moments by applying the method introduced in Hansen (1982).

Further, if intraday return over each interval (r_i) follows a normal distribution, $N(\mu, \sigma^2)$, $r_i - \mu$ should follow the distribution $N(0, \sigma^2)$. From the moment generating function, the moments of $r_i - \mu$ have the form of:

$$\begin{cases} E[(r_i - \mu)^{2n-1}] = 0 \\ E\left[(r_i - \mu)^{2n} - \dfrac{\sigma^{2n}(2n)!}{2^n n!}\right] = 0 \end{cases} \forall \text{ intergers } n \geq 1 \tag{2.22}$$

Therefore, we can form several moment conditions with the moment generation function. In our study, we conduct the test with four moment conditions.

$$g_T(\mu, \sigma^2) = \frac{1}{T}\sum_{i=1}^{T}\begin{pmatrix} r_i - \mu \\ (r_i - \mu)^2 - \sigma^2 \\ (r_i - \mu)^3 \\ (r_i - \mu)^4 - 3\sigma^4 \end{pmatrix} \tag{2.23}$$

The moment conditions are over-identified, and the analytical statistic follows the form:

$$Tg_T' {S_0}^{-1} g_T = \left(\frac{\left[\sum_{i=1}^{T}(r_i - \hat{\mu})^3\right]^2}{T6\hat{\sigma}^6} + \frac{\left[\sum_{i=1}^{T}(r_i - \hat{\mu})^4\right]^2}{T24\hat{\sigma}^8}\right) \tag{2.24}$$

Which follows a chi-square distribution with two degrees of freedom (χ^2_2).

2.3 Data

Our empirical study employs intraday trading data from the Thomson Reuters Tick History (TRTH) database from the SIRCA database. TRTH provides data for every single transaction record for individual stocks in U.S. stock markets. We obtained the relevant trade data for 30 Dow Jones stocks for the period from 1st January to 31st December 2012. These 30 stocks are listed in Appendix A. We report our results from GMM tests for the whole sample of firms during the sample period. Kernel estimation is conducted for each stock on each day, so we randomly choose to display the Kernel estimation results for the International Business Machines Corporation (IBM) stocks on 7th November 2012.

2.4 Empirical Results and Discussion

2.4.1 Correlation between Calendar Clock and Event Clock

We first investigate the frequency of the arrival of our events over traditional calendar clock intervals. Figure 2.1 shows the number of transactions and the amount of trading volume over 5-minute intervals and 10-minute intervals for IBM stock as at 7th November 2012. We witness a high correlation between the arrival of transactions and trading volume over calendar clock intervals. Our reason for using event clocks is to fulfil the condition to represent a constant level of market activity, so we find the transaction clock and the volume clock represent the level of market activity to a similar extent. On the other hand, if the calendar clock were to represent the level of market activity to the same extent as our event clocks, we would expect the arrival of transactions and trading volume to be stable through all times. In fact, the arrival of transactions and trading volume are dramatically different during the day, and the frequency of arrivals is much higher at the beginning and the end of trading hours. At this stage, we could conclude that the traditional calendar clock does not account for the level of market activity to measure time changes, while the transaction clock and the volume clock have the advantage of measuring time changes to represent the same level of market activity.

[Insert Figure 2.1]

2.4.2 Empirical Distribution with Kernel Estimator

We apply the Kernel method to estimate the empirical density of the intraday returns over calendar clock, transaction clock and volume clock. As expected, the stock returns over 5-minute intervals do not follow a normal distribution. More specifically, the empirical distribution in Figure 2.2 has a higher kurtosis than a normal distribution.

[Insert Figure 2.2]

As derived in section 2.2.2, the distribution of returns over calendar clock possibly follows a mixture of a normal distribution $\mathrm{N}(0, \beta\Delta t)$ and the scaled distribution of transactions $\left(\alpha - \frac{1}{2}\beta^2\right) n_i$ (or the scaled distribution of trading volume $\left(\alpha - \frac{1}{2}\beta^2\right) v_i$). Since the distribution of either transactions or trading volume over calendar clock is unknown, we again use the Kernel method to generate the empirical distribution, which is shown in Figure 2.3. We notice that transactions and trading volume are non-negative figures, and they both have a distribution positively skewed and left approaching zero. The mixture of normal distribution with a mean of 0, and a scaled, non-negative and positively skewed distribution would have a higher kurtosis and a positive skewness. This result is consistent with prior findings on the distribution of firm-level stock returns.

[Insert Figure 2.3]

Within the setting of transaction clock intervals, we observe a close-to-normal distribution of IBM stock returns over 100-transaction intervals in Figure 2.4a. Similarly, within the setting of volume clock intervals, a close to normal distribution can be observed for stock returns over 10,000-volume intervals in Figure 2.4b.

[Insert Figure 2.4]

Overall, we observe a close-to-normal distribution from the empirical distribution of returns over certain event clock intervals.

2.4.3 Empirical Results for Normality Tests

We perform the GMM tests with the analytical statistic on intraday returns of the Dow Jones 30 stocks over calendar clock, transaction clock and volume clock, respectively. The sample period contains all the 250 trading days in 2012. For the calendar clock, we set the intraday interval size to be five minutes, which is associated with 78 intervals per trading day; and ten minutes, which is associated with 39 intervals per trading day. For the transaction clock, we form two types of transaction clock intervals containing 100 transactions and 200 transactions, respectively. For the volume clock, we similarly form two types of volume clock intervals containing 20,000 shares and 40,000 shares to be traded, respectively.

Table 2.1 reports the results associated with the calendar clock intervals. With 5-minute intervals, for the Dow Jones 30 stocks, the percentage of days that the normality of intraday returns cannot be rejected ranges from 20% to 40.8%. With 10-minute intervals, the range increases to 43.6% to 60.8%. There seems to be insufficient evidence that intraday stock returns follow a normal distribution, as about half of the tests have rejected the null hypothesis of normality.

[Insert Table 2.1]

Table 2.2 reports the results associated with the transaction clock intervals. Among the Dow Jones 30 stocks, Cisco Systems Inc. (CSCO), Intel Corp (INTC) and Microsoft Corp (MSFT) seem to behave differently from the other stocks, as the average number of intervals for these three stocks turns out to be significantly larger than the other stocks. Excluding these three stocks, with 100-transaction intervals, the percentage of days that the normality of intraday returns cannot be rejected ranges from 62% to 77.2%. With 200-transaction intervals, the range increases to 72.4% to 86.8%. The choice of 100-transaction intervals and 200-transaction intervals provides a number of intraday intervals that is close to the number of 5-minute intervals and 10-minute intervals each day. We find that with a similar number of intraday intervals, the percentage of days with a normal distribution in intraday stock returns is higher when the transaction clock intervals are applied.

[Insert Table 2.2]

Table 2.3 reports the results associated with the volume clock intervals. Although we expected a high correlation between the number of transactions and the amount of trading volume, the results associated with the volume clock intervals hardly show a clear picture.

[Insert Table 2.3]

So far, we can tell that the transaction clock behaves better than the calendar clock and the volume clock in terms of observing normality of intraday returns distribution. In an attempt to fully explore the characteristics of transaction clocks, we conduct the same test with 20 different types of transaction clock. The size of the intervals ranges from 10 transactions per interval to 200 transactions per interval with an increment of 10 transactions each time. Tables 2.4 and 2.5 show that, excluding the three stocks mentioned before, with the first 10 types of intervals, percentage of days with normality ranges from 81.2% to 94.8%, while with all 20 types of intervals, the range becomes 92% to 99.6%.

[Insert Table 2.4 & 2.5]

Based on the empirical evidence from our tests, we find that stock returns do follow a normal distribution on an intraday basis when the transaction clock is applied.

2.5 Conclusion and Future Application

In this chapter, we introduce a different way to measure time changes using event clocks. We find that our proposed event clocks perform better than the traditional calendar clock in terms of capturing the level of market activity. Within the setting of the calendar clock, we find the distribution of returns follows a mixed distribution of a zero-mean normal distribution and a scaled, non-negative and positively skewed distribution of the number of transactions (or volume) in each interval. Within the settings of the event clock, we observe a close-to-normal empirical distribution of intraday stock returns from the Kernel density estimation. Our findings from GMM tests also suggest that intraday stock returns can follow a normal distribution when we use a transaction clock to measure time changes, which is consistent with Clark’s (1973) statement. We conclude that, when time changes are properly measured, there is a normal distribution of intraday stock returns.

Our findings suggest the transaction clock possesses good characteristics, such as capturing a constant level of market activity and uncovering a normal distribution of intraday stock returns. Applying the transaction clock will benefit future finance research in a range of areas, such as asset pricing, market microstructure and high-frequency trading studies on an intraday basis. The normality of stock returns fulfils the underlying assumptions of several asset pricing models, so testing those models with a transaction clock is likely to provide more meaningful results. Moreover, by capturing the same level of market activity, the transaction clock can benefit market microstructure studies by identifying abnormal transactions and root out insider trading activity. Further, given the existence of high-frequency trading, which allows dozens of transactions to take place in one single second, the transaction clock has an advantage over the traditional calendar clock in uncovering market behaviours and high-frequency trading strategies.

Figure 2.1a

Arrival of Transactions and Trading Volume over 5-minute Interval (IBM on 07/11/2012)

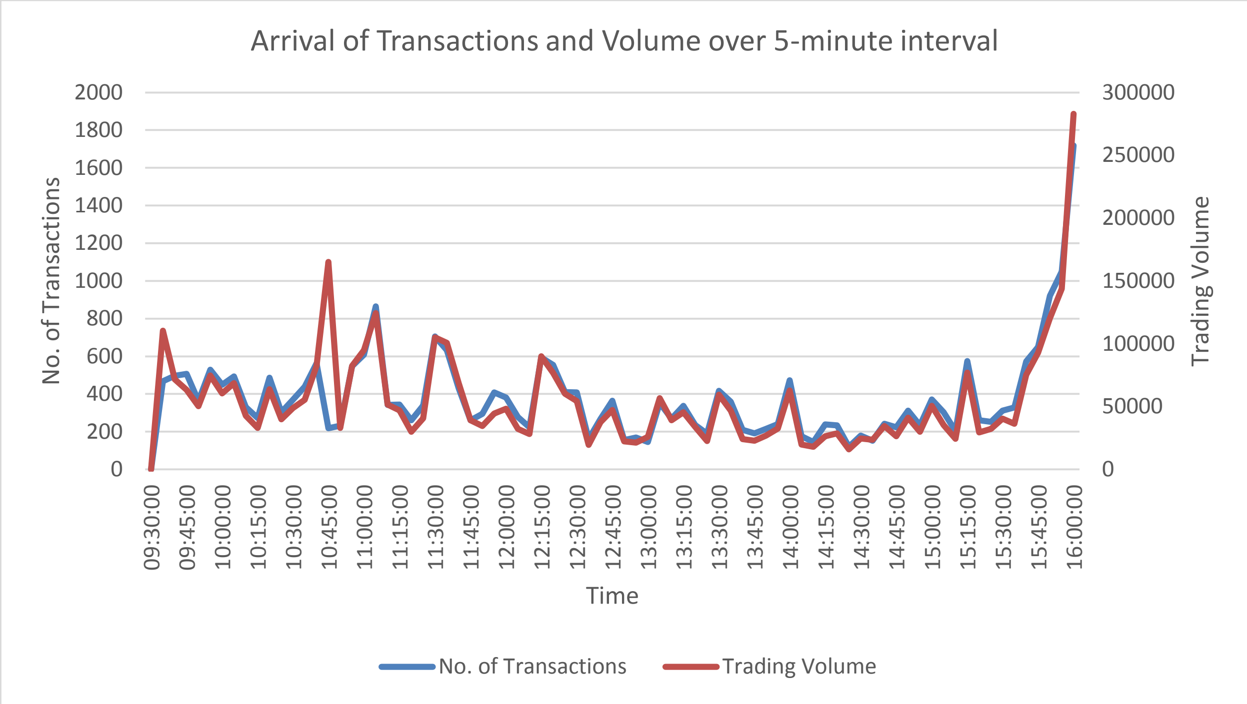

Figure 2.1b

Arrival of Transactions and Trading Volume over 10-minute Interval (IBM on 07/11/2012)

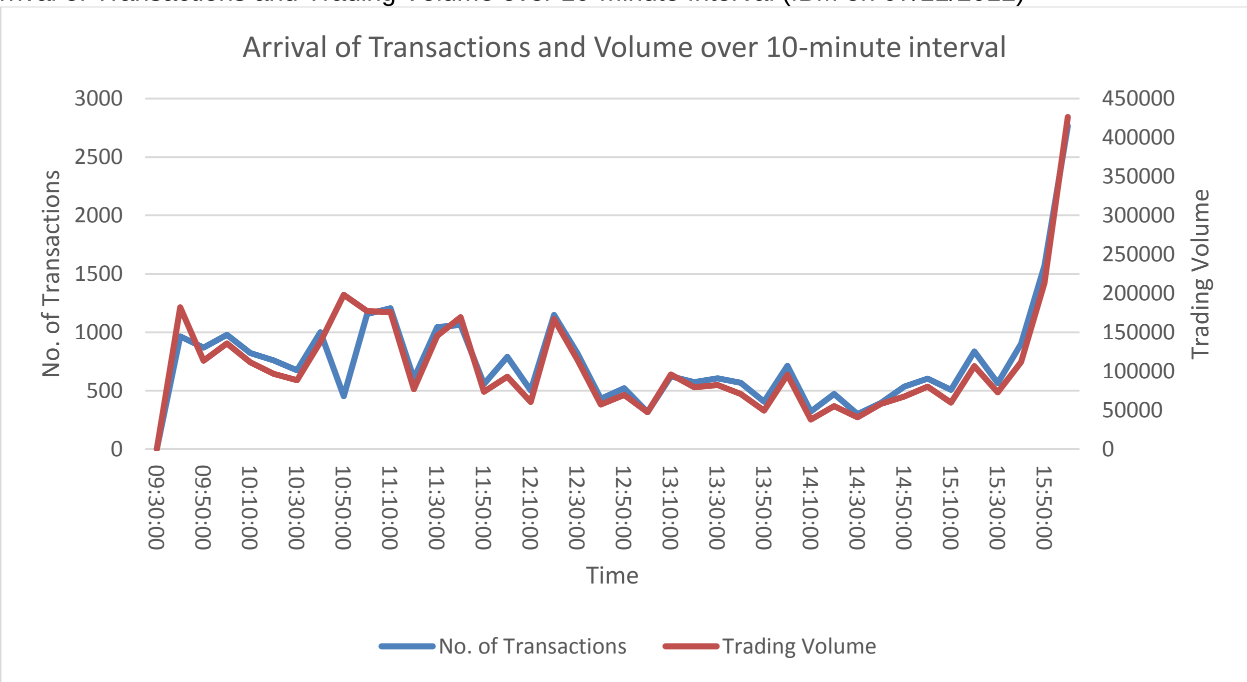

Figure 2.2

Empirical Distribution of Intraday Return over 5-minute Interval (IBM on 07/11/2012)

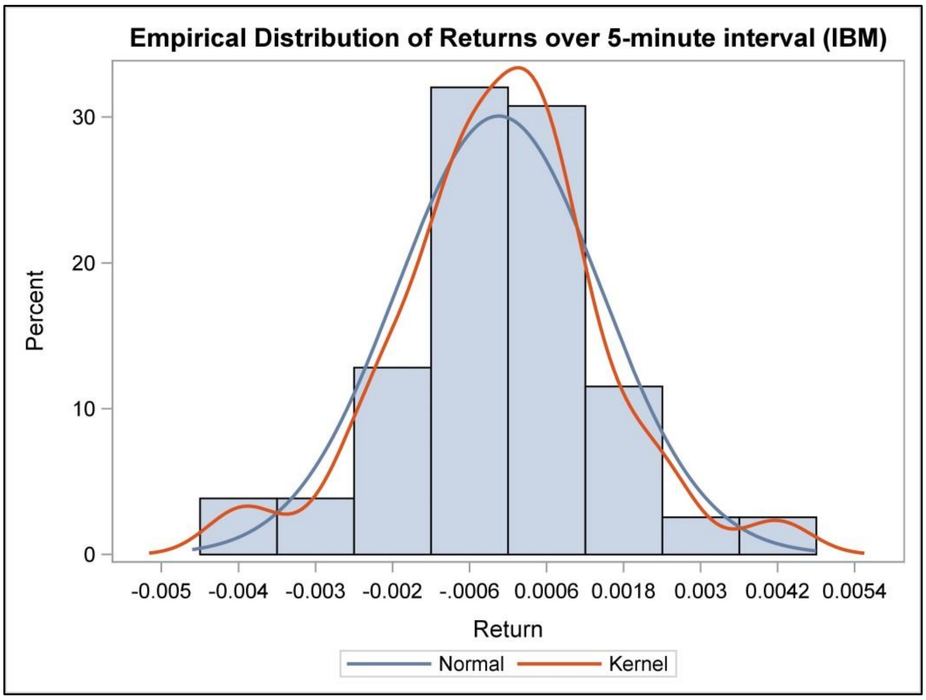

Figure 2.3a

Empirical Distribution of Transactions over 5-minute Interval (IBM on 07/11/2012)

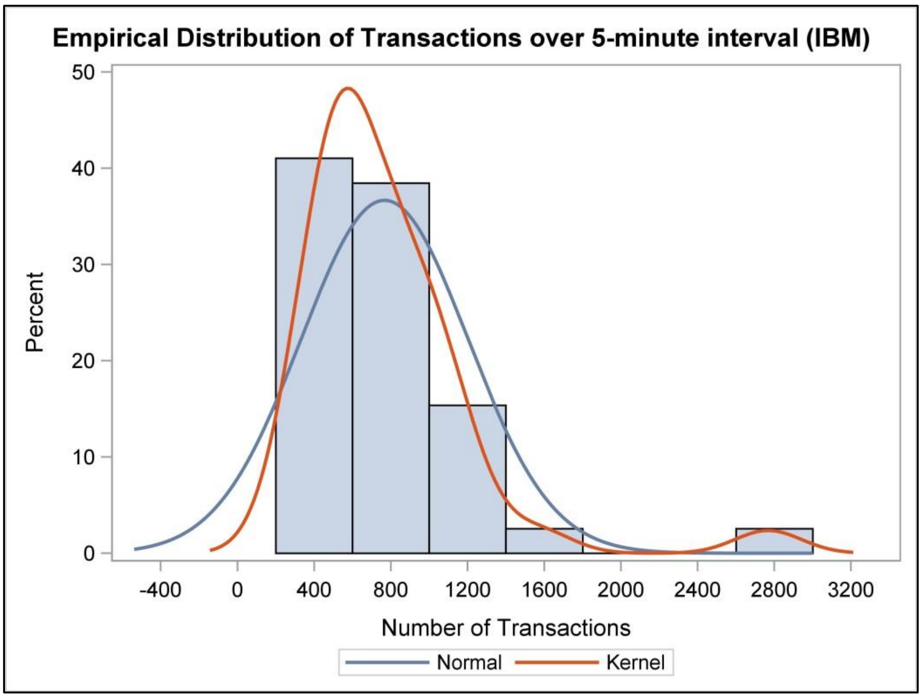

Figure 2.3b

Empirical Distribution of Trading Volume over 5-minute Interval (IBM on 07/11/2012)

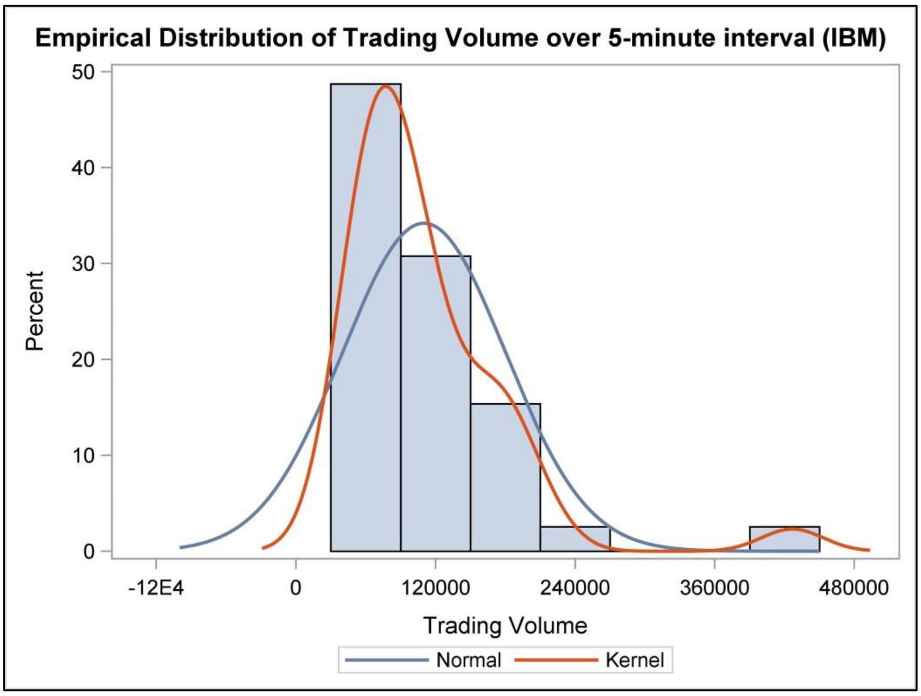

Figure 2.4a

Empirical Distribution of Intraday Return over 100-transaction Interval (IBM on 07/11/2012)

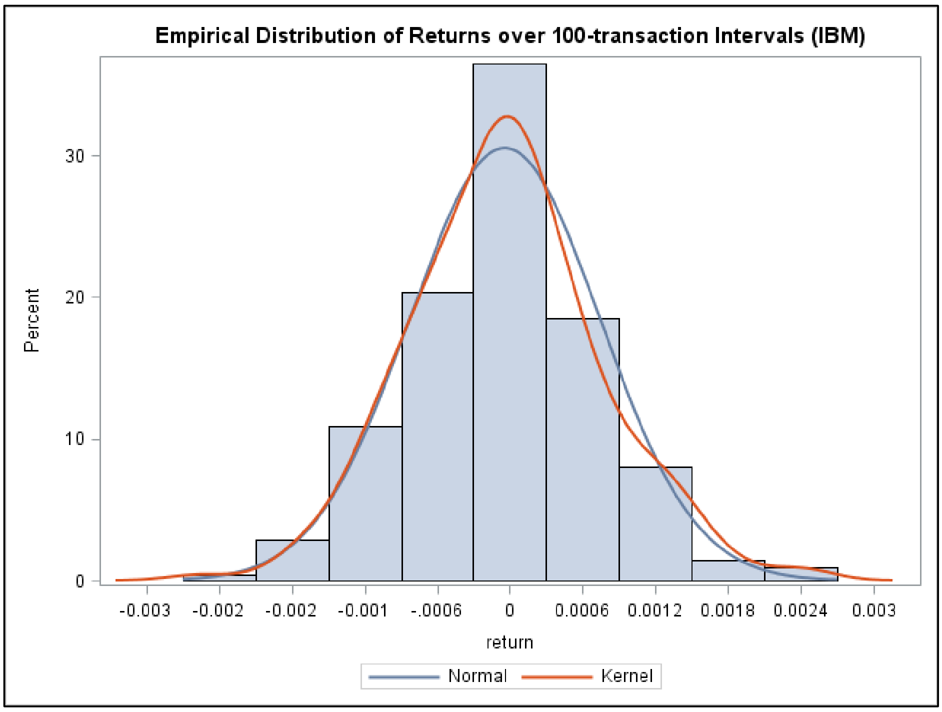

Figure 2.4b

Empirical Distribution of Intraday Return over 10,000-volume Interval (IBM on 07/11/2012)

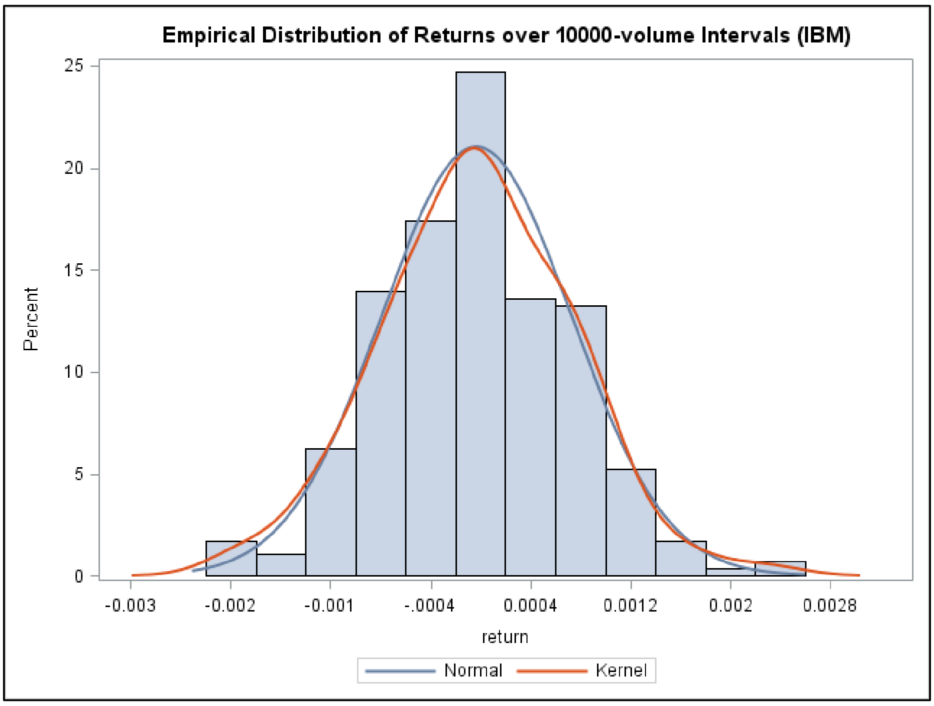

Table 2.1 Results of the GMM test on the normality of intraday stock returns distribution with a calendar clock

GMM test on Calendar Clock											
		5-minute interval		10-minute interval				5-minute interval		10-minute interval	
Co.	N	Average No. of intervals	%Normality	Average No. of intervals	%Normality	Co.	N	Average No. of intervals	%Normality	Average No. of intervals	%Normality
AXP	250	78	31.20%	39	52.80%	MCD	250	78	33.60%	39	51.60%
BA	250	78	27.20%	39	50.00%	MMM	250	78	30.80%	39	51.60%
CAT	250	78	25.20%	39	56.40%	MRK	250	78	37.60%	39	49.60%
CSCO	250	78	36.40%	39	52.80%	MSFT	250	78	29.60%	39	43.60%
CVX	250	78	37.20%	39	58.00%	NKE	250	78	29.60%	39	54.00%
DD	250	78	32.80%	39	60.80%	PFE	250	78	38.00%	39	57.20%
DIS	250	78	38.40%	39	58.80%	PG	250	78	33.60%	39	52.40%
GE	250	78	40.80%	39	60.40%	T	250	78	32.80%	39	51.20%
GS	250	78	24.40%	39	47.60%	TRV	250	78	29.20%	39	47.20%
HD	250	78	29.20%	39	53.60%	UNH	250	78	20.00%	39	50.40%
IBM	250	78	36.00%	39	54.40%	UTX	250	78	28.40%	39	47.20%
INTC	250	78	26.00%	39	49.20%	V	250	78	28.40%	39	56.40%
JNJ	250	78	32.40%	39	48.80%	VZ	250	78	37.60%	39	58.00%
JPM	250	78	27.20%	39	52.80%	WMT	250	78	30.00%	39	48.40%
KO	250	78	32.40%	39	53.20%	XOM	250	78	38.80%	39	59.20%

This table presents tests of normality of intraday stock returns over calendar clock intervals (5-minute intervals and 10-minute intervals) using GMM with the analytical statistic. The tests are performed using intraday observations on stock returns for the Dow Jones 30 firms over the sample period 1st January – 31st December 2012, containing 250 trading days in total. Columns 3, 5, 9 and 11 state the average number of intervals for a stock within a trading day. Columns 4, 6, 10 and 12 report the percentage of days that the normality of intraday returns cannot be rejected according to GMM tests.

Table 2.2 Results of the GMM test on the normality of intraday stock returns distribution with a transaction clock

GMM test on Transaction Clock											
		100-transaction interval		200-transaction interval				100-transaction interval		200-transaction interval	
Co.	N	Average No. of intervals	%Normality	Average No. of intervals	%Normality	Co.	N	Average No. of intervals	%Normality	Average No. of intervals	%Normality
AXP	250	42.18	66.00%	20.864	79.60%	MCD	250	48.48	65.20%	24.02	77.60%
BA	250	33.21	62.40%	16.368	79.60%	MMM	250	33.86	70.80%	16.672	80.80%
CAT	250	55.42	62.00%	27.448	78.40%	MRK	250	43.04	69.60%	21.284	82.40%
CSCO	250	970.48	0.00%	484.968	3.20%	MSFT	250	1290.51	0.00%	645.016	1.60%
CVX	250	60.7	64.40%	30.116	73.20%	NKE	250	34.32	58.80%	16.884	73.20%
DD	250	34.99	77.20%	17.268	86.80%	PFE	250	42.97	70.00%	21.256	81.60%
DIS	250	47.98	73.20%	23.748	80.80%	PG	250	46.73	66.00%	23.108	80.40%
GE	250	41.57	75.20%	20.568	82.80%	T	250	49.17	69.20%	24.316	78.80%
GS	250	51.04	65.20%	25.264	77.60%	TRV	250	25.95	69.60%	12.724	82.00%
HD	250	45.37	66.40%	22.44	80.40%	UNH	250	44.54	61.60%	22.024	72.40%
IBM	250	37.08	69.20%	18.32	78.00%	UTX	250	34.48	62.00%	16.976	80.00%
INTC	250	1125.54	0.40%	562.512	3.20%	V	250	30.85	64.80%	15.176	78.00%
JNJ	250	48.27	69.60%	23.896	83.20%	VZ	250	42.45	69.20%	20.988	80.40%
JPM	250	84.86	66.00%	42.18	78.00%	WMT	250	44.57	65.60%	22.04	81.60%
KO	250	48.18	70.00%	23.836	78.80%	XOM	250	89.83	62.40%	44.656	73.60%

This table presents tests of normality of intraday stock returns over transaction clock intervals (100-transaction interval and 200-transaction interval) using GMM with the analytical statistic. The tests are performed using intraday observations on stock returns for the Dow Jones 30 firms over the sample period 1st January – 31st December 2012, containing 250 trading days in total. Columns 3, 5, 9 and 11 state the average number of intervals for a stock within a trading day. Columns 4, 6, 10 and 12 report the percentage of days that the normality of intraday returns cannot be rejected according to GMM tests.

Table 2.3 Results of the GMM test on the normality of intraday stock returns distribution with a volume clock

GMM test on Volume Clock											
		20000-volume interval		40000-volume interval				20000-volume interval		40000-volume interval	
Co.	N	Average No. of intervals	%Normality	Average No. of intervals	%Normality	Co.	N	Average No. of intervals	%Normality	Average No. of intervals	%Normality
AXP	250	68.5	17.60%	33.77	44.40%	MCD	250	71.1	23.60%	35.07	45.60%
BA	250	46.49	25.60%	22.76	47.20%	MMM	250	41.06	38.00%	20.11	55.60%
CAT	250	69.14	26.00%	34.06	48.40%	MRK	250	126.67	4.00%	62.85	17.20%
CSCO	250	2088.56	0.00%	1044.2	0.00%	MSFT	250	2399.3	0.00%	1199.68	0.00%
CVX	250	85.59	13.60%	42.29	32.80%	NKE	250	33.25	41.20%	16.09	60.00%
DD	250	59.24	23.20%	29.08	54.40%	PFE	250	303.06	0.00%	151.04	4.00%
DIS	250	106.29	17.20%	52.63	35.60%	PG	250	111.97	6.80%	55.48	24.80%
GE	250	395.39	0.00%	197.22	4.00%	T	250	237.59	0.40%	118.3	9.20%
GS	250	54.96	36.40%	27	54.80%	TRV	250	35.58	29.60%	17.28	50.80%
HD	250	90.61	17.60%	44.83	33.20%	UNH	250	70.56	16.00%	34.8	42.80%
IBM	250	48.53	29.60%	23.72	52.80%	UTX	250	46.82	22.80%	22.9	41.20%
INTC	250	2117.67	0.00%	1058.95	0.00%	V	250	36.19	43.20%	17.71	62.80%
JNJ	250	126.77	4.00%	62.87	21.60%	VZ	250	128.28	6.00%	63.66	20.80%
JPM	250	280.23	1.60%	139.61	10.80%	WMT	250	95.24	11.20%	47.08	33.20%
KO	250	150.85	5.60%	74.94	22.80%	XOM	250	194.18	2.80%	96.61	12.00%

This table presents tests of normality of intraday stock returns over volume clock intervals (20,000-volume intervals and 40,000-volume intervals) using GMM with the analytical statistic. The tests are performed using intraday observations on stock returns for the Dow Jones 30 firms over the sample period 1st January – 31st December 2012, containing 250 trading days in total. Columns 3, 5, 9 and 11 state the average number of intervals for a stock within a trading day. Columns 4, 6, 10 and 12 report the percentage of days that the normality of intraday returns cannot be rejected according to GMM tests.

Table 2.4 Results on the normality of intraday stock returns with a range of 10 different transaction intervals

10-transaction interval to 100-transaction interval							
Co.	**Total days**	**Days with normality**	**%Normality**	**Co.**	**Total days**	**Days with normality**	**%Normality**
AXP	250	217	86.80%	MCD	250	229	91.60%
BA	250	219	87.60%	MMM	250	229	91.60%
CAT	250	225	90.00%	MRK	250	227	90.80%
CSCO	250	5	2.00%	MSFT	250	1	0.40%
CVX	250	214	85.60%	NKE	250	206	82.40%
DD	250	237	94.80%	PFE	250	233	93.20%
DIS	250	231	92.40%	PG	250	220	88.00%
GE	250	241	96.40%	T	250	231	92.40%
GS	250	215	86.00%	TRV	250	231	92.40%
HD	250	226	90.40%	UNH	250	203	81.20%
IBM	250	217	86.80%	UTX	250	214	85.60%
INTC	250	3	1.20%	V	250	212	84.80%
JNJ	250	231	92.40%	VZ	250	231	92.40%
JPM	250	226	90.40%	WMT	250	227	90.80%
KO	250	220	88.00%	XOM	250	222	88.80%

This table presents tests of normality of intraday stock returns over 10 different types of transaction clock intervals using GMM with the analytical statistic. The size of the intervals ranges from 10 transactions per interval to 100 transactions per interval, with an increment of 10 transactions each time. The tests are performed using intraday observations on stock returns for the Dow Jones 30 firms over the sample period 1st January – 31st December 2012, containing 250 trading days in total. Columns 3 and 6 state the number of days that at least with one type of transaction clock interval, the normality of intraday returns cannot be rejected according to GMM tests. Columns 4 and 8 provides the percentage of days that a normal distribution of intraday returns can be observed for each stock.

Table 2.5 Results on the normality of intraday stock returns with a range of 20 different transaction intervals

10-transaction interval to 200-transaction interval							
Co.	**Total days**	**Days with normality**	**%Normality**	**Co.**	**Total days**	**Days with normality**	**%Normality**
AXP	250	242	96.80%	MCD	250	242	96.80%
BA	250	241	96.40%	MMM	250	247	98.80%
CAT	250	245	98.00%	MRK	250	242	96.80%
CSCO	250	31	12.40%	MSFT	250	10	4.00%
CVX	250	242	96.80%	NKE	250	241	96.40%
DD	250	249	99.60%	PFE	250	244	97.60%
DIS	250	246	98.40%	PG	250	244	97.60%
GE	250	249	99.60%	T	250	246	98.40%
GS	250	242	96.80%	TRV	250	245	98.00%
HD	250	243	97.20%	UNH	250	230	92.00%
IBM	250	244	97.60%	UTX	250	243	97.20%
INTC	250	20	8.00%	V	250	240	96.00%
JNJ	250	248	99.20%	VZ	250	246	98.40%
JPM	250	241	96.40%	WMT	250	247	98.80%
KO	250	245	98.00%	XOM	250	244	97.60%

This table presents tests of normality of intraday stock returns over 20 different types of transaction clock intervals using GMM with the analytical statistic. The size of the intervals ranges from 10 transactions per interval to 200 transactions per interval, with an increment of 10 transactions each time. The tests are performed using intraday observations on stock returns for the Dow Jones 30 firms over the sample period 1st January – 31st December 2012, containing 250 trading days in total. Columns 3 and 6 state the number of days that at least with one type of transaction clock interval, the normality of intraday returns cannot be rejected according to GMM tests. Columns 4 and 8 provides the percentage of days that a normal distribution of intraday returns can be observed for each stock.

Chapter 3

INTRADAY BID-ASK SPREAD COMPONENTS AND PROBABILITY OF INFORMED TRADING

3.1 Introduction

The bid-ask spread of market makers, and its components, plays an important role in market microstructure studies. The bid-ask spread is regarded as cost of immediacy (Demsetz 1968); however, it also directly affects market makers' profits, investors' transaction costs and market liquidity. A high level of bid-ask spread might harm the market, because investors are not motivated to trade due to high transaction costs. On the other hand, a reasonably low level of bid-ask spread encourages investors to participate in the market and ensures a healthy trading environment. Therefore, it is important to analyse the bid-ask spread and its components, to better understand market design and the behaviour of market makers and investors. Moreover, since a market maker's bid-ask spread reflects his understanding of the level of information asymmetry for certain stocks, examining the bid-ask spread can reveal the impact of asymmetric information and to measure the degree of insider trading activity (Easley and O'Hara 1987; BSW 2004; Easley et al. 2012).

Prior research has made a significant contribution to understanding the determinants of the bid-ask spread. Stoll (1978) develops a theoretical model for bid-ask spread by identifying its three major cost components: order processing costs, inventory holding costs and adverse selection costs. Since then, several studies have explored proxy variables for the cost components, in order to explain the bid-ask spread. Bollen, Smith and Whaley (BSW) (2004) construct the inventory holding premium (IHP) as a proxy of inventory holding costs and adverse selection costs to model the market maker's bid-ask spread. Their model has high explanatory power for the spreads of a cross-section of stocks (Sidhu et al, 2008).

However, market conditions have changed significantly since the bid-ask spread literature was first developed. Over the last decade, the daily number of transactions, as well as the volume of market makers' bid and ask quotations has increased incredibly. Market activity has become far more frequent and transactions occur almost instantly. In this high-frequency market, it is better to use intraday observations than daily observations because intraday observations are recorded more frequently and capture more information during trading hours. Recent studies have analysed the intraday patterns of bid-ask spread since the 1990s (Chan et al. 1995; Li et al. 2005; Tannous et al. 2013; McInish and Wood 1992). O'Hara (2015) illustrates the impact of the current high-frequency market on market microstructure theories, and suggests that the classic models may be invalid under these conditions. Do market makers still exhibit the same behaviour in setting bid and ask quotations, and does the relation between the bid-ask spread and its determinants still hold? We believe it is vital to re-evaluate the determinants of the market maker's bid-ask spread in the current high-frequency market.

Adverse selection cost is a major cost component of the market maker's bid-ask spread; it has been widely used in market microstructure research as an indicator of the effect of asymmetric information. Adverse selection cost is defined as the cost that occurs when a market maker faces information asymmetry and is adversely selected by an informed trader. The relation between adverse selection costs and market maker's bid-ask spread allows us to quantify the impact of transactions initiated by informed traders. Moreover, studying bid-ask spreads in a high-frequency market on an intraday basis reveals the effect of asymmetric information in a short period (e.g. an hour, a day, and a week).

Our first objective is to re-evaluate the cost components of the market maker's bid-ask spread in the current high-frequency market, on an intraday basis. We employ the BSW bid-ask spread model and conduct empirical tests on frequently traded stocks in NYSE and NASDAQ. We chose the BSW model for two reasons. First, the variables in the model are less sensitive to significant increases in market activity than are most other proxies. Second, changing market conditions are not likely to violate the BSW's (2004) underlying assumption about market makers' behaviour in minimising inventory risk. Our findings suggest the BSW model has good explanatory power in the current high-frequency market.

Our second objective is to examine the potential benefits of applying the transaction clock in empirical tests. Chapter 2 illustrates that the transaction clock possesses the good characteristic of uncovering a normal distribution of intraday stock returns. Since the BSW model assumes a normal distribution of stock returns, we expect the tests in the transaction clock setting to provide more accurate and reliable results, as the distribution of intraday stock returns with a transaction clock fulfils the underlying assumption.

Our third objective is to introduce a method to estimate the probability of informed trading (PI) on a market level with intraday observations. We follow the method in the BSW model using an at-the-money (ATM) call option to measure a market maker's IHP as a proxy of inventory holding costs and adverse selection costs, and decompose it into the IHP of a transaction with an informed trader ($\mathrm{IHP_I}$) and the IHP of a transaction with an uninformed trader ($\mathrm{IHP_U}$). By examining the relation between the bid-ask spread and $\mathrm{IHP_I}$, we are able to estimate PI. Our method using intraday observations has advantages over the BSW (2004) method because it allows us to identify PI in a short period, such as an hour, a day and a week.

3.1.1 Bid-Ask Spread Components

Since Stoll (1978) first identified the costs components of the bid-ask spread, much effort has been devoted to finding good proxy variables for them. Potential components are now established in greater detail.

3.1.1.1 Order Processing Costs

Order processing costs are associated with market maker services. Such costs include the market makers' expenses on exchange seats, rents, bills, financial information, labour, etc. As these costs are likely to be fixed, their contribution to the market maker's bid-ask spread for each transaction should be related to his/her level of market activity. The more trading activity, the lower the contribution of these fixed costs to the spread of each transaction.

Trading volume is a widely used proxy to measure order processing costs. Trading volume in shares is used by Tinic (1972), Tinic and West (1972), Branch and Freed (1977) and BSW (2004). Trading volume in dollars is used in Stoll (1978) and Harris (1994). Harris (1994) also measures order processing costs by number of transactions. The findings in intraday studies on bid-ask spread components are consistent with the findings in prior studies on a daily basis. With intraday settings, order processing costs are associated with market activity, which is measured by number of transactions and trading volume (McInish and Van Ness, 2002).

3.1.1.2 Inventory Holding Costs and Adverse Selection Costs

Inventory holding costs occur when a market maker holds stocks to provide liquidity to the market. Such costs involve the opportunity cost of funds used to purchase and hold the stocks and potential stock value losses with adverse movements in stock prices. Prior literatures mostly use stock price (Demsetz 1968) and some volatility measures of stock returns (Tinic 1972; Tinic and West 1972; Stoll 1978; Harris 1994) to measure such costs. Adverse selection costs arise when a market maker is adversely selected by an informed trader. Several proxies have been constructed to measure adverse selection costs, such as concentration of insider ownerships (Glosten and Harris 1988), market capitalization (Harris 1994) and trading volume (Easley et al. 1996). BSW (2004) constructed the IHP to measure inventory holding costs and adverse selection costs as a whole. IHP allows market makers to minimise their risk relevant to these costs, and derive the payoff function of a market marker with minimal inventory risk to be the same as the payoff of an ATM call option with expiration given by the expected time the stock is held. Therefore, the IHP can be measured with the value of a hypothetical call option, which can be computed from Black and Scholes (1973) and Merton's (1973) option pricing model.

3.1.1.3 Competition Costs

Competition costs arise from competition among market makers. With certain competition, market makers are likely to adjust their bid-ask spread to make their quotes competitive and attractive to investors. As competition increases, the effect of the competition component on the spread will diminish. Tinic (1972) introduces the Herfindahl index of concentration to measure competition among market makers. BSW (2004) employ a modified version of Herfindahl index to measure the competition component of the bid-ask spread. Nowadays, the tremendous volume of market makers and the existence of dark pools mean that market makers face extremely high competition for order flows. Market makers tend to lower their bid-ask spread to attract order flows, rather than demand compensation from the bid-ask spread (Kwan et al. 2015). Therefore, the competition component is now expected to have little impact on market makers' bid-ask spreads.

3.1.2 Probability of Informed Trading

Asymmetric information has always been one of the key issues in market microstructure studies. Grossman (1976) established a model with independently and identically distributed private information to obtain rational expectation equilibrium. Within the model, the equilibrium fully reveals all private information. However, in reality, it is difficult to obtain fully revealing markets (Grossman,1967; Grossman and Stiglitz 1980). Admati (1985) extended Grossman's model and constructed a noisy rational expectation model for multi-assets in which the equilibrium does not necessarily fully reveal all the information.

Moreover, market participants with private information, who are categorised as informed traders, observe the true price of a stock and can benefit from their information by trading with others who do not possess the same information. From a market maker's perspective, a transaction initiated by an informed trader will result in a loss, which is the so-called adverse selection cost. Therefore, adverse selection cost becomes a good indicator for asymmetric information. Kyle (1985) has developed a model with a partially-revealing equilibrium when the market maker is adversely selected by informed traders. Admati and Pfleiderer (1988) and Foster and Viswanathan (1990) extend Kyle's study by letting the liquidity traders act strategically when informed traders start to trade on their information. Boulatov and George (2013) extend Kyle's model by including informed liquidity providers, and find that the existence of an informed liquidity provider results in wider bid-ask spreads, higher trading costs and lower market quality. Rosu (2015) models the equilibrium driven by the behaviour of informed and uninformed traders in order driven markets, and find that a higher share of informed traders reduces the bid-ask spread and improves liquidity.

Since market markers bear adverse selection costs, connections have been found between adverse selection costs and the market marker's bid-ask spread. Stoll (1978) identifies adverse selection costs as a major cost component of the bid-ask spread, which is supported by empirical evidence.

The established relation between adverse selection costs and the bid-ask spread provides a way to identify informed trading by analysing changes in bid-ask spreads. Easley and O'Hara (1987) analyse order imbalance from the bid-ask spread and introduce the probability of informed trading (PIN) as a measure of informed trading. Recently, they updated the original PIN measurement, and introduce the volume-based probability of informed trading (VPIN), which is indicated by order flows (Easley et al., 2012). Since VPIN is estimated in a high-frequency market, it is a way to measure the probability of informed trading in certain short periods.

BSW (2004) construct a bid-ask spread model with tick size, order processing costs, inventory holding costs, adverse selection costs and competition costs. They introduce IHP as a measure of adverse selection costs and inventory holding costs. The IHP of a transaction is modelled as an ATM call option with the expiration given by the expected time that the stock is held in inventory. In order to determine the amount of adverse selection costs, they decompose the IHP into two parts. The first part is the IHP of a transaction with an informed trader ($\mathrm{IHP_I}$), and the second part is the IHP of a transaction with an uninformed trader ($\mathrm{IHP_U}$). An informed trader observes the true price with a premium over the trading price, and the $\mathrm{IHP_I}$ can be modelled as an in-the-money (ITM) call option. An uninformed trader observes the quoted bid and ask prices in the market, and the $\mathrm{IHP_U}$ can be modelled as a slightly out-of-the-money (OTM) call option. The model works well with a cross-section of stocks on a daily basis, and estimates the PI for the market over a certain period (BSW 2004).

3.1.3 Intraday Event Clock Intervals

The current high-frequency market is characterised by an incredibly high volume of fast transactions. Daily observations may not be able to capture every piece of information during trading hours. On the other hand, intraday observations are recorded more frequently and are likely to contain more information than daily observations. It is critical to select an appropriate intraday setting because this directly affects the reliability and accuracy of the variable calculation as well as the statistical results. Most finance studies use a traditional calendar clock to measure time, in which each interval contains the same time length; in Chapter 2 we introduce a new time measurement using a transaction clock and find that it possesses good characteristics, such as capturing the same level of market activity and uncovering a normal distribution of intraday stock returns. The BSW model assumes a normal distribution of stock returns, which can be fulfilled by applying the transaction clock. Therefore, with the transaction clock, we expect to significantly improve the model in terms of better explanatory power and more reliable and accurate results.

3.2 Methodology

3.2.1 BSW Model

This study uses the BSW model for bid-ask spread. The market microstructure literature generally agrees that bid-ask spread is determined by market makers' order processing costs, competition costs, inventory holding costs and adverse selection costs. Following the literature, the BSW model uses trading volume to proxy order processing costs, Herfindahl index to proxy competition costs and inventory holding premium to proxy both inventory holding costs and adverse selection costs.

The model has the form of

$$\mathrm{SPRD_i} = \alpha_0 + \alpha_1 \mathrm{InvTV_i} + \alpha_2 \mathrm{MHI_i} + \alpha_3 \mathrm{IHP_i} + \varepsilon_i \tag{3.1}$$

where $\mathrm{SPRD_i}$ is the bid-ask spread of a stock in interval i, $\mathrm{InvTV_i}$ is the inverse of total trading volume in one interval, MHI is a modified version of the Herfindahl index, and IHP is the inventory holding premium which is the sum of inventory holding costs and adverse selection costs. The Herfindahl index is introduced by Tinic (1972) with the form:

$$\mathrm{HI} = \sum_{j=1}^{\mathrm{NM}} \left(\frac{\mathrm{V_j}}{\mathrm{TV}}\right)^2 \tag{3.2}$$

where $\mathrm{V_j}$ is the number of shares traded by market maker j, TV is the total number of shares, and NM is the number of market makers.

The inventory holding premium is modelled as an ATM call option that expires at the time the stock is held in inventory. Given the option valuation model developed by Black and Scholes (1973) and Merton (1973), the IHP follows the form:

$$\mathrm{IHP} = \mathrm{SN}\left(\frac{\ln\left(\frac{\mathrm{S}}{\mathrm{X}}\right)}{\sigma\sqrt{\mathrm{t}}} + 0.5\sigma\sqrt{\mathrm{t}}\right) - \mathrm{XN}\left(\frac{\ln\left(\frac{\mathrm{S}}{\mathrm{X}}\right)}{\sigma\sqrt{\mathrm{t}}} - 0.5\sigma\sqrt{\mathrm{t}}\right) \tag{3.3}$$

As for an ATM option, where S = X, it can be simplified to

$$\mathrm{IHP} = \mathrm{S}\left[2\mathrm{N}\left(0.5\sigma\mathrm{E}\left(\sqrt{\mathrm{t}}\right) - 1\right)\right] \tag{3.4}$$

where S is the stock price; σ is the annualized volatility of stock returns, $E(\sqrt{t})$ is the average square root of time between each transactions, and $N(\cdot)$ is the cumulative unit normal density function.

3.2.2 Modified Model with Intraday Settings

The original BSW model performs very well with a cross-section of shares on a daily basis. We modify the model to fit the intraday setting of the bid-ask spread from the literature. Moreover, we employ the transaction clocks introduced in Chapter 2, as well as traditional calendar clocks for intraday interval settings.

3.2.2.1 Order Processing Costs

According to Chapter 2, within intraday settings, the number of transactions better captures the level of market activity and is highly correlated with the amount of trading volume. We follow the BSW model and use trading volume as a proxy of order processing costs within the setting of calendar clock intervals. We also add the number of transactions as an alternative proxy to explore if it will improve the explanatory power of the model. Moreover, we use the transaction clock intervals from Chapter 2 as an alternative intraday interval setting, to test the BSW model. As each transaction clock interval contains the same number of transactions, there should be little difference in order processing costs in each interval. Therefore, within the settings of transaction clock intervals, we expect that order processing costs do not significantly affect the market maker's bid-ask spread.

3.2.2.2 Competition Costs

Intervals on an intraday basis are much narrower, in terms of time, number of transactions and trading volume. Not all market makers will get involved in the transactions in each interval. As the number of market makers varies in each intraday interval and is not recorded at such frequency, the Herfindahl index may not be a proper proxy for competition with intraday settings. Moreover, markets are now far more competitive and organized exchanges face intense competition from Dark trading pools. In highly competitive markets, market makers tend to lower their bid-ask spread to attract order flows, rather than demand compensation from the bid-ask spread (Kwan et al. 2015). The competition components should have little effect on the market maker's bid-ask spread. Therefore, we exclude the competition component from the BSW model in our study.

3.2.2.3 Inventory Holding Premium

We strictly follow the BSW model using IHP to represent inventory holding costs and adverse selection costs. As described in Chapter 2, stock returns follow a normal distribution when transaction clock intervals are set. IHP is modelled as an ATM call option, which can be valued with the Black and Scholes (1973) and Merton (1973) option pricing model. Since the underlying assumption of the option pricing model is stock return normality, we expect the model to perform better when transaction clock intervals are applied.

3.2.3 Informed vs Uniformed Traders

In the BSW model setting, IHP contains both inventory holding costs and adverse selection costs. As the adverse selection cost is the indicator of informed trading, a part of the IHP can be attributed to transactions with informed traders, while the rest can be attributed to transactions with uninformed traders. Therefore, the total IHP can be decomposed into the IHP of a transaction with an informed trader ($\mathrm{IHP_I}$), and the IHP of a transaction with an uninformed trader ($\mathrm{IHP_U}$).

$$\mathrm{IHP} = \mathrm{PI} \times \mathrm{IHP_I} + (1 - \mathrm{PI}) \times \mathrm{IHP_U} \tag{3.5}$$

where PI represents the probability of informed trading.

The difference between informed and uninformed traders is that informed traders are able to observe the true price of a stock while uninformed traders can only observe market prices. Therefore, the IHP for both types of traders can still be modelled using a call option, with the exercise price to be the ask price, where the 'stock price' is the price observed by each type of trader.

Uninformed traders observe the market price of a stock, which is indicated by the latest bid and ask prices. The $\mathrm{IHP_U}$ can be treated as a slightly OTM call option with the stock price to be the mid-quote price and the exercise price to be the ask price.

On the other hand, informed traders, who have better information, are able to observe the true price of a stock. The $\mathrm{IHP_I}$ can be treated as an ITM call option with the stock price to be the observed true price and the exercise price to be the ask price. However, the true price is not observable. Following the setting in the original paper, the true price is allowed a premium over the exercise price by 1% to 10%. For simplicity, we set the premium to be 5%.

Therefore $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$ take the form:

$$\mathrm{IHP_k} = \mathrm{S_k N}\left(\frac{\ln\left(\frac{\mathrm{S_k}}{\mathrm{X}}\right)}{\sigma\sqrt{\mathrm{t}}} + 0.5\sigma\sqrt{\mathrm{t}}\right) - \mathrm{XN}\left(\frac{\ln\left(\frac{\mathrm{S_k}}{\mathrm{X}}\right)}{\sigma\sqrt{\mathrm{t}}} - 0.5\sigma\sqrt{\mathrm{t}}\right) \tag{3.6}$$

where k=I or U, X = ask price, S_U = mid-quote price, and S_I = ask price * (1+5%).

According to BSW model, the relation between bid-ask spread and IHP should follow:

$$\mathrm{SPRD} = \alpha_0 + \alpha_1 \mathrm{IHP} + \varepsilon \tag{3.7}$$

Substituting IHP with the components for informed and uniformed traders,

$$\mathrm{SPRD} = \alpha_0 + \alpha_1(\mathrm{PI} \times \mathrm{IHP_I} + (1 - \mathrm{PI}) \times \mathrm{IHP_U}) + \varepsilon \tag{3.8}$$

Rearranging the model above, it has the form:

$$\mathrm{SPRD} = \alpha_0 + \alpha_1 \mathrm{IHP_U} + \alpha_1 \times \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon \tag{3.9}$$

By restricting α_1= 1, the model becomes:

$$\mathrm{SPRD} = \alpha_0 + \mathrm{IHP_U} + \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon \tag{3.10}$$

The coefficient on $(\mathrm{IHP_I} - \mathrm{IHP_U})$ is PI, which can be estimated by model above.

3.3 Data

The intraday trade and quotation data are collected from the Thomson Reuter Tick History (TRTH) from SIRCA database. We obtained the relevant trade and quote data for Dow Jones 30 stocks, of which 27 stocks are listed in NYSE and 3 stocks are listed in NASDAQ. A list of sample stocks is presented in Appendix A. The sample period is from 1st January to 31st December 2012. The data contains all the transaction records with price and volume, and the entire bid and ask prices quoted by the market makers. Since not all quotes will result in a transaction, the number of quotations significantly exceeds the number of transactions in the same trading day. To find the quotation associated with a particular transaction, we match a transaction with the quote immediately prior to that transaction.

3.4 Empirical Results and Discussion

3.4.1 Summary Statistics

With the matched dataset, we compute two types of spreads that are generally used in prior studies.

The first one is the quoted spread:

$$\text{Quoted Spread}_{\mathrm{i}} = \text{Ask price}_{\mathrm{i}} - \text{Bid price}_{\mathrm{i}} \tag{3.11}$$

where i represents the i-th transaction of a particular stock. Based on the quoted spread of each transaction, we compute an equally-weighted average of quoted spreads (EWQS) to represent the spread over an intraday interval.

The second one is the effective spread:

$$\text{Effective Spread}_{\mathrm{i}} = 2|\text{Trade price}_{\mathrm{i}} - \text{Midquote}_{\mathrm{i}}| \tag{3.12}$$

where

$$\text{Midquote}_{\mathrm{i}} = \frac{(\text{Bid price}_{\mathrm{i}} + \text{Ask price}_{\mathrm{i}})}{2} \tag{3.13}$$

With the effective spread of each trade, we compute the volume-weighted average of effective spreads (VWES) to represent the spread over an intraday interval.

Stock price (S) is the average trade price in each interval. InvTV is simply calculated as the inverse of total trading volume in each interval. NT is the number of transaction in each interval. $\overline{\sqrt{\mathrm{t}}}$ is the average of square root of time between transactions in each interval. σ is the annualized realised volatility of log returns in each interval.

Table 3.1 presents the summary statistics of the variables across DJ30 stocks during the sample period. Panel A contains the statistics using intraday calendar clock intervals (5-minute intervals) and Panel B contains the statistics using intraday transaction clock intervals (100-transaction intervals). With calendar clock intervals, equal-weighed quoted spread (EWQS) and value weighted effective spread (VWES) have an average of $0.0190 and $0.0149, respectively. The average price (S) is $66.94 across the sample stocks. The number of shares traded (TV) has an average of 58,137 shares and a median of 9,471 shares, while the number of transactions (NT) has an average of 168 transactions and a median of 38 transactions. The significant difference between the average and the median of TV and NT can be attributed to high market trading activity shortly after a market opens and before it closes.

Comparing Panel A and B, the number of observations using transaction clock intervals is almost twice the number of observations with calendar clock intervals. As the transaction clock intervals arrive more frequently, it is likely to capture smaller variable fluctuations, more often. With the same number of transactions in each interval, market activity is distributed evenly across the intervals. We also find that the values of all four spread measures are stable with transaction clock intervals, while the values fluctuate a lot with calendar clock intervals. Similar results can be observed on spread determinants, such as trading volume, annualised stock return volatility and IHP. The stable variable values can be attributed to the characteristic of the transaction clock in capturing a constant level of market activity. Moreover, we observe lower values in variables such as EWQS, VWES, S, $\sqrt{t}$ and IHP in Panel B, indicating that stocks with lower prices and spreads are traded more frequently.

[Insert Table 3.1]

Table 3.2 contains the estimates of the correlations between variables. Panel A provides the cross-correlation between variables with intraday calendar clock intervals. As expected, the correlation between EWQS and VWES is 0.9591, indicating a high correlation between two spread measures. High correlations between the independent variables (InvTV, InvNT and IHP) and the spread measures (EWQS and VWES) are consistent with our expectation, suggesting InvTV and IHP are closely related to the bid-ask spread. Moreover, the correlation between TV and NT is 0.9433, indicating the two measurements of the market activity are highly correlated. In Panel B, we observe similar results with the transaction clock intervals.

[Insert Table 3.2]

3.4.2 BSW Model with Intraday Intervals

We first test the original BSW model and the modified model with intraday observations of a cross-section of stocks. Table 3.3 provides the results for the regressions on equal-weighted quoted spread. Panel A and B contain the results with calendar clock intervals and transaction clock intervals, respectively. Since InvTV, InvNT and IHP are modelled as the cost components of the bid-ask spread, we expect the coefficients on these variables to be significantly positive. We observe significantly positive coefficients on IHP in the original and modified BSW model, indicating significant impact of inventory costs and adverse selection costs on the quoted bid-ask spread. The values of the coefficients are consistently stable across all the models with both calendar clock intervals and transaction clock intervals. However, we fail to observe significant positive coefficients on InvTV and InvNT, which suggests limited impact of order processing costs on the quoted bid-ask spread. The inconsistent coefficients on InvTV and InvNT are contrary to our expectations; nevertheless, the original BSW model has a R-square of 0.6537, indicating that the cost components (InvTV and IHP) explain more than 65% of the bid-ask spread. Replacing trading volume with number of transactions slightly improves the explanatory power, as the R-square increases to 0.6608. Including both trading volume and number of transactions in the model brings the R-square to 0.6624, indicating that trading volume does not provide additional explanatory power when the model contains number of transactions. The number of transactions better represents order processing costs on an intraday basis, albeit only slightly. Moreover, within the setting of transaction clock intervals, the models are significantly improved with a higher R-square of 0.7136, and the coefficient on InvTV is consistent with our expectation. By excluding the order processing cost (InvTV) from the model, with transaction clock intervals, we still observe an R-square of 0.7126, which is not significantly different from the model including such variable. Therefore, while inventory costs and adverse selection costs are still significant cost components of the bid-ask spread, order processing costs seem to be not as significant.

[Insert Table 3.3]

Table 3.4 provides the results for the regressions on volume-weighted effective spread. With calendar clock intervals, we observe similar results to those for EWQS. However, the correlation between IHP and VWES is almost half the correlation between IHP and EWQS, indicating that IHP has a higher impact on quoted spread than effective spread. Moreover, the R-square is slightly lower, which suggests relatively lower explanatory power on VWES. With the transaction clock, the model improves with a higher R-square, which is consistent with our findings on EWQS.

[Insert table 3.4]

We then perform the same test on individual stocks. Table 3.5 presents the results for the explanatory power of the original BSW model within the settings of the calendar clock and the transaction clock. Panel A reports the R-squares of the model with the EWQS as the spread measure. The results are consistent with the ones on a cross-section of stocks. The R-square is over 0.3 for most of the stocks with the calendar clock, indicating that the model possesses good explanatory power on the bid-ask spread of individual stocks. When the transaction clock is applied, the R-square increases by approximately 0.1 for most of the stocks, which is also consistent with our findings on the cross-section of stocks. Similar findings can be observed in Panel B, and the R-squares are relatively lower when the VWES is the spread measure. Overall, in an intraday setting for individual stocks, quoted spreads outperform effective spreads, while transaction clock intervals outperform traditional calendar clock intervals in terms of increasing the model's explanatory power.

[Insert table 3.5]

3.4.3 Probability of Informed Trading (PI) Estimation

We estimate the PI for a cross-section of stocks over certain periods using an intraday setting. Table 3.6 presents the PI estimation and the associated t-stat with the settings of the calendar clock and the transaction clock. Panel A, B, C and D provide the PI estimation over 1-year period, 1-week period, 1-day period and 1-hour period respectively. We randomly choose the 1-week period to be the week starting 1st October 2012, the 1-day period to be 1st October 2012 (Monday) and the 1-hour period to be the first trading hour (9:30-10:30) on 1st October 2012. Despite the different time length in each period, the model works well in PI estimation with high R-squares and reasonable statistical significance in the coefficients. We also observe higher R-squares with the transaction clock intervals, and the shorter the period, the more R-square increases. Although the periods are randomly selected, PIs in the 1-year period, 1-week period and 1-day period remain stable, which are consistently around 0.6%, while PI in the first trading hour of a day appears to be significantly higher. Here we merely provide some examples of PI estimation; further tests can be conducted on PI in the periods of interest.

[Insert table 3.6]

3.5 Conclusion

This chapter re-evaluates the determinants of a market maker's bid-ask spread in a high-frequency market on an intraday basis. Our findings suggest that the BSW model possesses good explanatory power, despite significantly changed market conditions. With more frequent posting of bid and ask quotations and of delivering transactions, market makers behave rationally in setting their bid-ask spreads. While order processing costs seem not to have significant impact on bid-ask spread, inventory costs and adverse selection costs still significantly affect the bid-ask spread in a high-frequency market. As a result, market makers demand significant compensation from the bid-ask spread for these types of costs. We also find that quoted spread is a better spread measure than effective spread, in order to improve the explanatory power of the model.

For comparison purposes, our empirical tests of the BSW model apply the transaction clock as well as the traditional calendar clock. Unlike the calendar clock, the transaction clock uncovers a normal distribution of intraday stock returns, which fulfils the model's underlying assumption. The coefficients on the independent variables are consistent across all the models with both clocks; however, the transaction clock outperforms the traditional calendar clock in terms of explanatory power. Our findings on the transaction clock reveal the potential benefits of applying the transaction clock in financial studies, especially under an assumption of return distribution normality.

Moreover, we develop a method to estimate the probability of informed trading (PI) on a market level, with intraday observations. Our proposed method has advantages over the BSW (2004) method for identifying PI over certain short periods, such as an hour, a day and a week. This allows us to more accurately identify the impact of asymmetric information and root out insider trading activity.

Table 3.1 Summary statistics with a calendar clock and a transaction clock

This table presents the summary of descriptive statistics of variables for DJ30 stocks in intraday calendar clock intervals and transaction clock intervals. EWQS is the equal-weighted quoted bid-ask spread; VWES is the volume-weighted effective bid-ask spread; REWQS is EWQS divided by stock price; RVWES is VWES divided by stock price; S is the average stock price in each interval; InvS is the inverse of S; TV is the number of shares traded in each interval; InvTV is the inverse of TV; NT is the number of transactions in each interval; σ is the annualized realized volatility of stock return in each interval; $\overline{\sqrt{t}}$ is the average of the square root of time between transactions; and IHP is the inventory holding premium. The sample period is from 1st January to 31st December 2012.

	Calendar Clock (5-minute intervals)					Transaction Clock (100-transaction intervals)				
Variables	nobs	mean	25%	median	75%	nobs	mean	25%	median	75%
Spread measures										
EWQS	550060	0.0190	0.0103	0.0125	0.0202	920044	0.0120	0.0100	0.0100	0.0100
VWES	550060	0.0149	0.0098	0.0106	0.0152	920044	0.0102	0.0088	0.0096	0.0100
REWQS	550060	0.0003	0.0002	0.0003	0.0004	920044	0.0004	0.0003	0.0004	0.0005
RVWES	550060	0.0002	0.0002	0.0002	0.0003	920044	0.0003	0.0003	0.0003	0.0004
Determinants of spread										
S	550060	66.94	38.83	61.35	87.25	920044	35.12	20.56	27.42	31.82
InvS	550060	0.0205	0.0115	0.0163	0.0257	920044	0.0367	0.0314	0.0365	0.0486
TV	550060	55147	4900	9650	23397	920044	35023	23621	30198	39473
InvTV	550060	0.0002	0.0000	0.0001	0.0002	920044	0.0000	0.0000	0.0000	0.0000
NT	550060	160	24	38	65	920044	100	100	100	100
$\sqrt{t}$	550060	0.1733	0.0861	0.1236	0.1923	920044	0.9163	0.2127	0.4239	0.6655
σ	550060	0.0009	0.0006	0.0009	0.0012	920044	0.0002	0.0000	0.0001	0.0003
IHP	550060	0.0033	0.0013	0.0023	0.0041	920044	0.0011	0.0003	0.0005	0.0010

Table 3.2 Summary of cross-correlations with a calendar clock and a transaction clock

This table presents the cross-correlations between variables for DJ30 stocks in intraday calendar clock intervals (Panel A) and transaction clock intervals (Panel B). EWQS is the equal-weighted quoted bid-ask spread; VWES is the volume-weighted effective bid-ask spread; REWQS is EWQS divided by stock price; RVWES is VWES divided by stock price; S is the average stock price in each interval; $\mathrm{Inv}S$ is the inverse of S; TV is the number of shares traded in each interval; InvTV is the inverse of TV; NT is the number of transactions in each interval; σ is the annualized realized volatility of stock return in each interval; $\overline{\sqrt{t}}$ is the average of the square root of time between transactions; and IHP is the inventory holding premium. The sample period is from 1st January to 31st December 2012.

Panel A. Calendar clock (5-minute intervals) (n=5550060)											
Variables	EWQS	VWES	REWQS	RVWES	S	InvS	TV	InvTV	NT	$\sqrt{t}$	σ
VWES	0.9591										
REWQS	0.4791	0.4420									
RVWES	0.2260	0.2813	0.8958								
S	0.7084	0.6919	-0.1812	-0.3872							
InvS	-0.4467	-0.4271	0.4686	0.6863	-0.7908						
TV	-0.1737	-0.1695	0.2646	0.3576	-0.3424	0.5273					
InvTV	0.3243	0.3278	-0.0601	-0.1644	0.4180	-0.4131	-0.2615				
NT	-0.1661	-0.1646	0.2516	0.3360	-0.3333	0.5093	0.9433	-0.2667			
$\sqrt{t}$	0.1922	0.2087	-0.2050	-0.2643	0.3472	-0.4474	-0.5003	0.7606	-0.5400		
σ	0.0533	0.0318	0.5780	0.5798	-0.3171	0.5416	0.7202	-0.3025	0.7659	-0.5854	
IHP	0.8061	0.7809	0.2453	0.0227	0.7000	-0.5488	-0.2792	0.4705	-0.2787	0.4013	0.0020

Panel B. Transaction clock (100-transaction intervals) (n=920044)											
Variable	EWQS	VWES	REWQS	RVWES	S	InvS	TV	InvTV	NT	$\sqrt{t}$	σ
VWES	0.9576										
REWQS	0.0898	0.0471									
RVWES	-0.1025	-0.0214	0.8653								
S	0.7172	0.7057	-0.4979	-0.5904							
InvS	-0.4537	-0.4498	0.7840	0.8327	-0.8187						
TV	-0.1618	-0.1256	0.1328	0.2152	-0.2120	0.2095					
InvTV	0.4498	0.3919	-0.1753	-0.3283	0.4923	-0.4151	-0.6443				
NT											
$\sqrt{t}$	0.4394	0.4309	-0.4196	-0.5041	0.7014	-0.6692	-0.0535	0.3263			
σ	-0.0302	-0.0265	0.0866	0.0969	-0.0696	0.0978	0.0280	-0.0596		-0.0938	
IHP	0.8441	0.8040	-0.1686	-0.3391	0.8149	-0.6186	-0.1775	0.4993		0.7093	-0.0577

Table 3.3 Regression results of equal-weighted quoted spread

This table presents the results of the cross-sectional regression results of equal-weighted quote spread of DJ30 stocks using the original BSW model and the modified model. EWQS is the equal-weighted quoted spread; InvTV is the inverse of number of shares traded; InvNT is the inverse of number of transactions; and IHP is the inventory holding premium. The sample period is from 1st January to 31st December 2012.

Panel A contains the results of 3 regressions with calendar clock intervals:

Model 1: $\mathrm{EWQS} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_3 \mathrm{IHP} + \varepsilon$

Model 2: $\mathrm{EWQS} = \alpha_0 + \alpha_2 \mathrm{InvNT} + \alpha_3 \mathrm{IHP} + \varepsilon$

Model 3: $\mathrm{EWQS} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_2 \mathrm{InvNT} + \alpha_3 \mathrm{IHP} + \varepsilon$

Panel B contains the results of 2 regressions with transaction clock intervals:

Model 4: $\mathrm{EWQS} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_3 \mathrm{IHP} + \varepsilon$

Model 5: $\mathrm{EWQS} = \alpha_0 + \alpha_3 \mathrm{IHP} + \varepsilon$

Panel A. Calendar clock (5-minute intervals)						
		Coefficient estimates / t-stat				
EWQS	Nobs	Intercept	InvTV	InvNT	IHP	R^2
Model 1	550060	0.0056	-6.0877		4.4050	0.6537
		287.3318	-78.4681		933.3530	
Model 2	550060	0.0067		-0.0643	4.4076	0.6608
		308.7494		-133.9101	1018.4711	
Model 3	550060	0.0070	6.5086	-0.0972	4.3120	0.6624
		309.7912	49.8355	-119.0817	912.6522	

Panel B. Transaction clock (100-transaction intervals)						
		Coefficient estimates / t-stat				
EWQS	Nobs	Intercept	InvTV		IHP	R^2
Model 4	920044	0.0069	22.7015		4.0353	0.7136
		532.1505	58.5584		1281.6479	
Model 5	920044	0.0076			4.1273	0.7126
		1398.1329			1510.2153	

Table 3.4 Regression results of volume-weighted effective spread

This table presents the results of the cross-sectional regression results of volume-weighted effective spread of DJ30 stocks using the original BSW model and the modified model. VWQS is the volume-weighted effective spread; InvTV is the inverse of number of shares traded; and IHP is the inventory holding premium. The sample period is from 1st January to 31st December 2012.

Panel A contains the results of 3 regressions with calendar clock intervals:

Model 1: $\mathrm{VWES} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_3 \mathrm{IHP} + \varepsilon$

Model 2: $\mathrm{VWES} = \alpha_0 + \alpha_2 \mathrm{InvNT} + \alpha_3 \mathrm{IHP} + \varepsilon$

Model 3: $\mathrm{VWES} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_2 \mathrm{InvNT} + \alpha_3 \mathrm{IHP} + \varepsilon$

Panel B contains the results of 2 regressions with transaction clock intervals:

Model 4: $\mathrm{VWES} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_3 \mathrm{IHP} + \varepsilon$

Model 5: $\mathrm{VWES} = \alpha_0 + \alpha_3 \mathrm{IHP} + \varepsilon$

Panel A. Calendar clock (5-minute intervals)						
		Coefficient estimates / t-stat				
VWES	Nobs	Intercept	InvTV	InvNT	IHP	R^2
Model 1	550060	0.0062	-2.9096		2.8018	0.6117
		453.2231	-53.3969		845.2557	
Model 2	550060	0.0067		-0.0300	2.8011	0.6152
		438.1676		-88.4664	916.0445	
Model 3	550060	0.0068	2.8408	-0.0444	2.7594	0.6159
		426.7808	30.7417	-76.8307	825.4150	

Panel B. Transaction clock (100-transaction intervals)						
		Coefficient estimates / t-stat				
VWES	Nobs	Intercept	InvTV		IHP	R^2
Model 4	920044	0.0077	-4.9419		2.5540	0.6466
		822.9439	-17.8013		1132.7701	
Model 5	920044	0.0075			2.5340	0.6464
		1930.2602			1296.9594	

Table 3.5 The model's explanatory power on individual stocks with a calendar clock and a transaction clock

This table presents results for the test on the explanatory power of the original BSW model on individual stocks within the intraday settings of the calendar clock and the transaction clock. SPRD is the spread measure; InvTV is the inverse of number of shares traded; and IHP is the inventory holding premium. Panel A employs EWQS as the spread measure, and Panel B employs VWES as the spread measure. The Dow Jones 30 stocks are chosen as the sample stocks. The sample period is from 1st January to 31st December 2012.

Model:

$$\mathrm{SPRD} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_3 \mathrm{IHP} + \varepsilon$$

Panel A					
	R^2			R^2	
SYMBOL	Calendar clock	Transaction Clock	SYMBOL	Calendar clock	Transaction clock
AXP	0.4128	0.4786	MCD	0.3788	0.4897
BA	0.3611	0.4961	MMM	0.4127	0.5504
CAT	0.3888	0.5425	MRK	0.2001	0.3068
CSCO	0.1750	0.0092	MSFT	0.4088	0.0130
CVX	0.3971	0.4918	NKE	0.4557	0.6383
DD	0.3197	0.4287	PFE	0.0035	0.0245
DIS	0.3128	0.4642	PG	0.1730	0.3623
GE	0.0016	0.0056	T	0.0819	0.1864
GS	0.3621	0.5205	TRV	0.4264	0.6048
HD	0.2952	0.4168	UNH	0.4221	0.5270
IBM	0.4091	0.5882	UTX	0.4111	0.5793
INTC	0.1189	0.0108	V	0.4048	0.6002
JNJ	0.2373	0.3981	VZ	0.1873	0.3370
JPM	0.2006	0.1752	WMT	0.3375	0.4564
KO	0.3374	0.4462	XOM	0.3166	0.4074

Panel B					
	R^2			R^2	
SYMBOL	Calendar clock	Transaction Clock	SYMBOL	Calendar clock	Transaction clock
AXP	0.3513	0.3845	MCD	0.3133	0.4125
BA	0.3075	0.4496	MMM	0.3496	0.5034
CAT	0.3262	0.4855	MRK	0.1680	0.2218
CSCO	0.2055	0.2228	MSFT	0.1054	0.1662
CVX	0.3270	0.4248	NKE	0.4206	0.6111
DD	0.2679	0.3954	PFE	0.0160	0.0418
DIS	0.2527	0.3692	PG	0.1975	0.2871
GE	0.0077	0.0096	T	0.0658	0.0756
GS	0.3132	0.4857	TRV	0.3888	0.5629
HD	0.2153	0.2797	UNH	0.3626	0.4702
IBM	0.3414	0.5423	UTX	0.3554	0.5377
INTC	0.0942	0.1974	V	0.3500	0.5697
JNJ	0.1557	0.2290	VZ	0.1526	0.2215
JPM	0.1104	0.1313	WMT	0.2670	0.3944
KO	0.2553	0.3008	XOM	0.2237	0.2746

Table 3.6 Estimation of probability of informed trading in certain periods

This table presents results for the PI estimation of the DJ30 stocks over certain periods. SPRD is the spread measure, which is equal-weighted quoted spread (EWQS); $\mathrm{IHP_I}$ is the inventory holding premium of a transaction with an informed trader; $\mathrm{IHP_U}$ is the inventory holding premium of a transaction with an uninformed trader; and PI is the probability of informed for a stock during the period. Panel A provides PI estimation over the whole 1-year sample period from 1st January to 31st December 2012; Panel B provides PI estimation over an 1-week period from 1st October to 5th October 2012; Panel C provides PI estimation over an 1-day period on 1st October 2012; and Panel D provides PI estimation over an 1-hour period from 09:30am to 10:30am on 1st October 2012.

Model:

$$\mathrm{SPRD} = \alpha_0 + \alpha_1 \mathrm{IHP_U} + \alpha_1 \times \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon$$

with the restriction of $\alpha_1 = 1$.

Panel A. 1-year period (01/01/2012-31/12/2012)

Calendar clock (5-minute intervals)

EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I} - \mathrm{IHP_U}$	R^2
		Coefficient estimates / t-stat			
	550060	-0.0014	1.00	0.0059	0.5165
		-45.3864		718.2935	

Transaction clock (100-transaction intervals)

EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I} - \mathrm{IHP_U}$	R^2
		Coefficient estimates / t-stat			
	920044	0.0041	1.00	0.0044	0.5313
		388.7528		924.9788	

Panel B. 1-week period (01/10/2012-05/10/2012)

Calendar clock (5-minute intervals)

EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I} - \mathrm{IHP_U}$	R^2
		Coefficient estimates / t-stat			
	11089	-0.0020	1.00	0.0060	0.5153
		-8.6972		102.4883	

Transaction clock (100-transaction intervals)

EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I} - \mathrm{IHP_U}$	R^2
		Coefficient estimates / t-stat			
	17960	0.0038	1.00	0.0047	0.5588
		52.5572		137.7421	

Panel C. 1-day period (01/10/2012)					
Calendar clock (5-minute intervals)					
		Coefficient estimates / t-stat			
EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I - IHP_U}$	$\mathrm{R^2}$
	2219	-0.0006	1.00	0.0058	0.5037
		-1.0614		45.0915	

Transaction clock (100-transaction intervals)					
		Coefficient estimates / t-stat			
EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I - IHP_U}$	$\mathrm{R^2}$
	4120	0.0031	1.00	0.0053	0.5924
		18.8770		71.0366	

Panel D. 1-hour period (09:30-10:30, 01/10/2012)					
Calendar clock (5-minute intervals)					
		Coefficient estimates / t-stat			
EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I - IHP_U}$	$\mathrm{R^2}$
	318	-0.0001	1.00	0.0088	0.4103
		-0.0531		14.4994	

Transaction clock (100-transaction intervals)					
		Coefficient estimates / t-stat			
EWQS	Nobs	Intercept	$\mathrm{IHP_U}$	$\mathrm{IHP_I - IHP_U}$	$\mathrm{R^2}$
	866	-0.0009	1.00	0.0089	0.6603
		-1.7114		38.5571	

Chapter 4

ADVERSE SELECTION IN THE GLOBAL FINANCIAL CRISIS

4.1 Introduction

The 2007–2009 global financial crisis (GFC) seriously affected world financial systems and led to severe crashes in many stock markets. For instance, the Dow Jones Industrial Average lost more than 50% of its value during the crisis period. Since the fundamental value of the stocks did not necessarily drop by that much during the crisis, the crashes have mainly been attributed to investor overreaction. History shows that stock markets crash during financial crises and recover afterwards, but it seems that investors never learn from the past, and continue to overreact to crisis. What makes investors act this way?

Morris and Shin (2012) argue that stock market crashes can be attributed to investors losing confidence and worrying about being adversely selected by informed traders during a financial crisis. These authors believe that markets are confident when participants commonly understand the fundamental soundness of the market, and only traders with market confidence participate. In a healthy trading environment, uninformed traders participate in the market because they believe the trades are mutually beneficial, despite experiencing loss when they are adversely affected by informed traders. When a shock arrives, an uninformed trader may be concerned that other uninformed traders will exit and leave him/her to trade with informed traders. Thus, he/she will have less confidence in the market and be less motivated to trade. Consequently, as uninformed traders exit, informed traders make up a greater portion of market participants, resulting in higher adverse selection during the crisis. Therefore, comparing adverse selection during a crisis period and a non-crisis period will help us better understand why investors overreact during a crisis.

The market microstructure literature has established that adverse selection occurs when a market participant is adversely selected by an informed trader, which is a consequence of information asymmetry. Since it is impossible to identify all the market participants involved in transactions with informed traders, prior studies have tended to focus on a specific group of participants, the market makers. Market makers provide liquidity to the market, and therefore trade with both uninformed and informed traders. From a market maker's perspective, transactions initiated by informed traders cause them a loss, because informed traders have better information. Following this logic, several studies have attempted to model adverse selection through market makers behaviours. Kyle (1985) develops a model with a partially-revealing equilibrium when a market maker is adversely selected by informed traders. To extend Kyle's study, Admati and Pfleiderer (1988) and Foster and Viswanathan (1990) let the liquidity traders act strategically when informed traders start to trade on their information. Boulatov and George (2013) further extend Kyle's study by including informed liquidity providers, and find that the existence of an informed liquidity provider increases the bid-ask spreads and trading costs, and reduces market quality.

When a market maker trades with an informed trader, he will bear the cost of adverse selection. To cover this and other costs, the market maker carefully sets his bid-ask spread by placing bid and ask quotes. Not surprisingly, prior studies have found tight connections between adverse selection costs and the market maker's bid-ask spread. Empirical evidence supports the argument that adverse selection costs are a major cost component of the bid-ask spread (Stoll 1978). Therefore, we are able to identify informed trading from the relation between adverse selection costs and bid-ask spreads. Easley and O'Hara (1987) introduce a method to estimate the probability of informed trading (PIN) by analysing the order imbalance of bid-ask spreads. The volume-based probability of informed trading (VPIN), which is indicated by order flows, is proposed to be a more accurate measure in a high-frequency market (Easley et al., 2012).

BSW (2004) construct a bid-ask spread model with tick size, order processing costs, inventory holding costs, adverse selection costs and competition costs. They model adverse selection cost and inventory holding cost together as the inventory holding premium (IHP). The IHP is modelled as an at-the-money (ATM) call option with expiration given by the expected time that the stock is held in inventory. For each transaction which could be initiated by an informed trader, the market maker's IHP can be decomposed into two parts for informed and uninformed trading, respectively. As the informed traders observe the true price with a premium over the trading price, the IHP for informed traders can be modelled as an in-the-money (ITM) call option. Similarly, the IHP for uninformed traders can be modelled as a slightly out-of-the-money (OTM) call option. The model works well with a cross-section of stocks on a daily basis, and provides estimates of probability of informed trading (PI) for the market in certain periods (BSW 2004).

It has taken decades to establish a fundamental understanding of bid-ask spread and adverse selection. In the meantime, market conditions have changed significantly, with or without the impact of the GFC. During the last decade, we have witnessed a tremendous increase in daily transactions, trading volume and volume of bid and ask quotations placed by market makers. O'Hara (2015) illustrates the impact of the current high-frequency market on market microstructure theories, and challenges the validity of the previous bid-ask spread models. In Chapter 3, we re-evaluate the cost components of the market makers' bid-ask spread in the current high-frequency market within an intraday setting, and find that the BSW model has good explanatory power. Zhu and Gippel (2015) also apply the BSW (2004) method to estimate weekly PI to examine information asymmetry on debt covenant violations. In this chapter, we employ the BSW model as our base model because it consistently performs well under changing market conditions.

Our first objective is to investigate the impact of the financial crisis on the stock market and on market makers' behaviours. We use the BSW bid-ask spread model to perform empirical tests on a cross-section of stocks during crisis and non-crisis periods. Our second objective is to apply the method of PI estimation from Chapter 3 to examine the impact of the GFC on adverse selection in stock trading on a market level. Our third objective is to introduce a new method to estimate the daily PI for individual stocks, and apply the method to examine adverse selection on individual stocks during crisis and non-crisis periods. We follow the method in BSW (2004) to decompose a market maker's IHP into the IHP of a transaction with an informed trader ($\mathrm{IHP_I}$), and the IHP of a transaction with an uninformed trader ($\mathrm{IHP_U}$). With a chance to be adversely selected by an informed trader, the expected IHP for a transaction on one stock is the probability weighted average of $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. We are able to estimate the PIs of individual stocks by solving a fully identified equation, which is constructed with the relation between IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$,. By examining the PIs of the market and individual stocks, we find a significantly higher level of adverse selection during a crisis period relative to a non-crisis period, at both a market level and an individual stock level.

The remainder of this chapter is structured as follows: Section 4.2 introduces the methods we use to examine the impact of the GFC and construct the short-term PI estimation for individual stocks; Section 4.3 describes the data; Section 4.4 presents and discusses the empirical results and Section 4.5 concludes the chapter.

4.2 Methodology

4.2.1 BSW Model

We employ the BSW model for bid-ask spreads as a base model in this study. The market microstructure literature has established that the bid-ask spread is commonly modelled as the costs of market makers. Its cost components include order processing costs, inventory holding costs and adverse selection costs. The BSW model contains these components, and uses an ATM call option to model the IHP, which accounts for both inventory holding costs and adverse selection costs.

The basic form of the BSW model is

$$\mathrm{SPRD}_{i,t} = \alpha_0 + \alpha_1 \mathrm{InvTV}_{i,t} + \alpha_2 \mathrm{IHP}_{i,t} + \varepsilon_{i,t} \tag{4.1}$$

where $\mathrm{SPRD}_{i,t}$ is the bid-ask spread of stock i in interval t, InvTV is the inverse of total trading volume in each interval, and IHP is the inventory holding premium.

The IHP is modelled as an ATM call option with expiration given by the expected time that the stock is held in inventory. Given the option valuation model developed by Black and Scholes (1973) and Merton (1973), the IHP follows the form:

$$\mathrm{IHP} = \mathrm{SN}\left(\frac{\ln\left(\frac{S}{X}\right)}{\sigma\sqrt{t}} + 0.5\sigma\sqrt{t}\right) - \mathrm{XN}\left(\frac{\ln\left(\frac{S}{X}\right)}{\sigma\sqrt{t}} - 0.5\sigma\sqrt{t}\right) \tag{4.2}$$

where S is the stock price; σ is the annualized volatility of stock returns, $E(\sqrt{t})$ is the average square root of time between each transactions, and $N(\cdot)$ is the cumulative unit normal density function.

As for an ATM option, where S = X, it can be simplified to

$$\mathrm{IHP} = S\left[2N\left(0.5\sigma E\left(\sqrt{t}\right) - 1\right)\right] \tag{4.3}$$

4.2.2 Intraday Interval Setting

According to our findings in Chapter 3, the BSW model has good explanatory power with a cross-section of shares on a daily basis (BSW 2004) as well as on an intraday basis. In this study, we perform the test on a cross-section of shares on an intraday basis. We employ a traditional calendar time clock, and set the intraday interval to be a fixed 10-minute. The first interval of a trading day contains observations in the first 10 minutes after the market opens, and the second interval contains observations in the following 10 minutes. This setting provides 39 constant intervals in a trading day.

4.2.3 Modified Model with Crisis Period Dummies

To uncover the difference in adverse selection during the crisis and non-crisis period, we employ two dummy variables to account for the period before the crisis and the period after the crisis respectively. The model with period dummy variables has the following form:

$$\begin{aligned} \mathrm{SPRD}_{i,t} = {} & \alpha_0 + \alpha_1 \mathrm{InvTV}_{i,t} + \alpha_2 \mathrm{InvTV}_{i,t} \times d_1 + \alpha_3 \mathrm{InvTV}_{i,t} \times d_2 \\ & + \alpha_4 \mathrm{IHP}_{i,t} + \alpha_5 \mathrm{IHP}_{i,t} \times d_1 + \alpha_6 \mathrm{IHP}_{i,t} \times d_2 + \varepsilon_{i,t} \end{aligned} \tag{4.4}$$

where $\mathrm{SPRD}_{i,t}$ is the bid-ask spread of a stock i in interval t, InvTV is the inverse of total trading volume in one interval, IHP is the inventory holding premium which is the sum of inventory holding costs and adverse selection costs, d_1 is the pre-GFC period dummy, and d_2 is the post-GFC period dummy.

4.2.4 Informed vs Uniformed Traders

We strictly follow the BSW model using the IHP to represent inventory holding costs and adverse selection costs.

In the BSW model setting, the IHP contains both inventory holding costs and adverse selection costs. As the adverse selection cost is the indicator of informed trading, a part of the IHP can be attributed to transactions with informed traders, while the rest can be attributed to transactions with uninformed traders. We set the IHP for a transaction with an informed trader to be IHP_I, and the IHP for a transaction with an uninformed trader to be IHP_U. For a single transaction which may be adversely selected by an informed trader, a market maker's IHP would be the probability-weighted average of $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. Therefore, the total IHP can be decomposed into $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$:

$$\mathrm{IHP} = \mathrm{PI} \times \mathrm{IHP_I} + (1 - \mathrm{PI}) \times \mathrm{IHP_U} \tag{4.5}$$

where PI represents the probability of informed trading.

The difference between informed and uninformed traders is that informed traders are able to observe the true price of a stock while uninformed traders can only observe the market price. Therefore, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$ can be modelled using a call option with the exercise price set to be the ask price, where the 'stock price' is the price observed by each type of trader.

Uninformed traders observe the market price of a stock, which is indicated by the latest bid and ask prices. The $\mathrm{IHP_U}$ can be treated as a slightly OTM call option with the stock price to be the mid-quote price and the exercise price to be the ask price.

On the other hand, informed traders, who have better information, are able to observe the true price of a stock. The $\mathrm{IHP_I}$ can be treated as an ITM call option with the stock price to be the observed true price and the exercise price to be the ask price. However, the true price is not observable. Following the setting in BSW (2004), the true price is allowed a premium over the exercise price of 1% to 10%. For simplicity, we set the premium to be 5%. Therefore $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$ take the form:

$$\mathrm{IHP_k} = \mathrm{S_k} \mathrm{N}\left(\frac{\ln\left(\frac{\mathrm{S_k}}{\mathrm{X}}\right)}{\sigma\sqrt{\mathrm{t}}} + 0.5\sigma\sqrt{\mathrm{t}}\right) - \mathrm{XN}\left(\frac{\ln\left(\frac{\mathrm{S_k}}{\mathrm{X}}\right)}{\sigma\sqrt{\mathrm{t}}} - 0.5\sigma\sqrt{\mathrm{t}}\right) \tag{4.6}$$

where k=I or U, X = ask price, S_U = mid-quote price, and S_I = ask price * (1+5%).

Applying the Black, Scholes (1973) and Merton (1973) option pricing model allows us to quantify the IHP for different types of traders, and thus estimate the PI.

First, we follow the method in the original paper to use a restricted regression to estimate PI of a cross-section of stocks. According to the BSW model, the relation between bid-ask spread and IHP should follow:

$$\mathrm{SPRD} = \alpha_0 + \alpha_1 \mathrm{IHP} + \varepsilon \tag{4.7}$$

Substituting IHP with the components for informed and uniformed traders,

$$\mathrm{SPRD} = \alpha_0 + \alpha_1(\mathrm{PI} \times \mathrm{IHP_I} + (1 - \mathrm{PI}) \times \mathrm{IHP_U}) + \varepsilon \tag{4.8}$$

Rearranging the model above, it has the form:

$$\mathrm{SPRD} = \alpha_0 + \alpha_1\mathrm{IHP_U} + \alpha_1 \times \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon \tag{4.9}$$

By restricting α_1= 1, the model becomes:

$$\mathrm{SPRD} = \alpha_0 + \mathrm{IHP_U} + \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon \tag{4.10}$$

The coefficient on $(\mathrm{IHP_I} - \mathrm{IHP_U})$ is PI, which can be estimated by the model above.

Second, we introduce a new method to estimate daily PIs for individual stocks. On a single transaction, a market maker considers the chance of being adversely selected by an informed trader, and his bid-ask spread reflects his understanding of the probability of informed trading. The relation between a market maker's bid-ask spread and his understanding of the probability of informed trading at a time is described in Equation (4.8). Equation (4.8) is derived by replacing IHP with its equivalent in Equation (4.5), which rewrites IHP as the probability weighted average of $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. Therefore we can directly infer a PI from Equation (4.5) for each transaction. Similarly, for a period with a number of transactions, we can infer a PI from Equation (4.5) using accumulated IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$, which accounts for the probability of informed trading for the period.

To estimate the daily PI for an individual stock, we set the period to be a single trading day, and calculate the equal-weighted average and the volume-weighted average for IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. We construct the equation with the following form:

$$\mathrm{EWIHP} = \mathrm{PI} \times \mathrm{EWIHP_I} + (1 - \mathrm{PI}) \times \mathrm{EWIHP_U} \tag{4.11}$$

and

$$\mathrm{VWIHP} = \mathrm{PI} \times \mathrm{VWIHP_I} + (1 - \mathrm{PI}) \times \mathrm{VWIHP_U} \tag{4.12}$$

where EWIHP, $\mathrm{EWIHP_I}$ and $\mathrm{EWIHP_U}$ are equal-weighted average IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$ in one trading day respectively, and VWIHP, $\mathrm{VWIHP_I}$ and $\mathrm{VWIHP_U}$ are volume-weighted average IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$ in one trading day respectively.

Both equations are fully identified with one unknown (PI), which can be easily solved, and PI indicates the daily probability of informed trading for one stock.

The daily PI of individual stocks allows us to examine the impact of the GFC on adverse selection in the stock market for individual stocks. With the estimated daily PI of our sample stocks, we perform a simple test with the GFC-period dummy:

$$\mathrm{PI} = \alpha_0 + \alpha_1 \mathrm{d_{GFC}} + \varepsilon \qquad (4.13)$$

where PI is the daily probability of informed trading of an individual stock and $\mathrm{d_{GFC}}$is the GFC-period dummy.

4.3 Data

We collect the intraday trade and quotation data from Trade and Quotation (TAQ) and 30-day realized volatility data from Option Metrics. Both databases are located at Wharton Research Data Services (WRDS). The sample period covers the pre-GFC period from 1st July 2005 to 30th June 2007, GFC period from 1st July 2007 to 30th June 2009, and post-GFC period from 1st July 2009 to 30th June 2011. For reporting purposes, we report the results for August 2005 as a pre-GFC period, August 2007 as a GFC period, and August 2009 as a post-GFC period. We select our sample stocks from the Dow Jones 30. The components of Dow Jones 30 stocks altered four times during our sample period, so for consistency, we choose the same 25 stock components through the whole period. Appendix B lists the 25 sample stocks. We obtain the relevant trade and quote data for these 25 stocks, including all transaction records with price and volume, and all bid and ask prices quoted by market makers. Since not all quotes result in transactions, the number of quotations significantly exceeds the number of transactions in each trading day. To find the quotation associated with a particular transaction, we match a transaction with the quote immediately prior to that transaction.

4.4 Empirical Results and Discussion

4.4.1 Summary Statistics

With the matched dataset, we compute two types of spreads that are generally used in prior studies.

The first one is the quoted spread:

$$\text{Quoted Spread}_{\mathrm{i}} = \text{Ask price}_{\mathrm{i}} - \text{Bid price}_{\mathrm{i}} \tag{4.14}$$

where i represents the i-th transaction of a particular stock. Based on the quoted spread of each transaction, we compute an equal-weighted average of quoted spreads (EWQS) to represent the spread over an intraday interval.

The second one is the effective spread:

$$\text{Effective Spread}_{\mathrm{i}} = 2|\text{Trade price}_{\mathrm{i}} - \text{Midquote}_{\mathrm{i}}| \tag{4.15}$$

where

$$\text{Midquote}_{\mathrm{i}} = \frac{(\text{Bid price}_{\mathrm{i}} + \text{Ask price}_{\mathrm{i}})}{2} \tag{4.16}$$

With the effective spread of each trade, we compute the volume-weighted average of effective spreads (VWES) to represent the spread over an intraday interval.

Stock price (S) is the average trade price in each interval. InvTV is simply calculated as the inverse of total trading volume in each interval. NT is the number of transactions in each interval. $\overline{\sqrt{\mathrm{t}}}$ is the average of square root of time between transactions in each interval. σ is the annualized realized volatility of log returns in each interval.

Table 4.1 presents the summary statistics of the variables across 25 sample stocks during the sample period. We employ the traditional calendar clock to set the intraday intervals. The intervals are equally spaced with a constant 10-minute time length. Panel A contains the statistics of variables during pre-GFC period (August 2005). Panel B contains the statistics during GFC period (August 2007). Panel C contains the statistics during post-GFC period (August 2009). We find the bid-ask spreads are a lot wider during the crisis period than they are during the non-crisis periods. EWQS has an average of 0.1538 during the GFC period and an average of 0.0921 and 0.0249 during the pre-GFC period and post-GFC period, respectively. Similar results can be observed on other spread measures. This result is not surprising. On the one hand, market makers tend to be more conservative during a crisis period, and demand more from the spread to cover their costs. On the other hand, the number of sell orders is likely to largely exceed the number of buy orders, and such imbalance would also widen the spreads.

Not surprisingly, the average 30-day realized volatility significantly increased from 0.1726 to 0.2750 during the GFC period, and slightly reduced to 0.2704 after the crisis. This is the result of the dramatic price changes that occur during market crash and market recovery. Moreover, the average inventory holding premium also increased from 0.0023 to 0.0038 during the crisis period and reduced to 0.0027 after the crisis, indicating that market makes bore much higher inventory and adverse selection costs during the GFC.

[Insert Table 4.1]

Interestingly, average trading volume decreased during the GFC period and the post-GFC period, despite greater numbers of transactions during both periods. It is understandable that transactions were more frequent, but the decrease in trading volume suggests that the transactions also became smaller. Moreover, in Table 4.2, we find the correlation between trading volume and the number of transactions to be 0.7075 during the pre-GFC period, 0.4970 during the GFC period and 0.6251 during the post-GFC period. In Chapter 3, we observed the correlation between trading volume and number of transactions to be close to 1. Apparently, a correlation of 0.4970 during the GFC period is abnormal, and can be attributed to the impact of the crisis.

[Insert Table 4.2]

Table 4.2 contains the estimates of the correlation between the variables. Panels A, B and C provide the cross-correlation between variables during the pre-GFC period, GFC period and post-GFC period, respectively. EWQS and VWES are the independent variables in our regression model and are shown to have high correlations across the whole sample period. We also find that correlations between spread measures (EWQS and VWES) and spread determinants (InvTV and IHP) remain consistent across all three periods, despite the impact of the crisis.

4.4.2 Adverse Selection from Cross-Sectional Evidence

We first test the modified BSW model with intraday observations of a cross-section of stocks. The model is described in section 4.2.3. The results are reported in Table 4.3. Panel A provides the results for the two regressions on EWQS, and Panel B provides the results for the two regressions on VWES. Models 1 and 3 contain variables including InvTV, IHP and their interaction with non-crisis period dummies, while models 2 and 4 only contain variables including IHP and its interaction with non-crisis period dummies. As cost components of the bid-ask spread, InvTV and IHP are expected to be positively correlated with the spread measures. Moreover, comparing the coefficients of InvTV and IHP with the coefficients of their interactions with period dummies allows us to analyse market makers' behaviours during the crisis and non-crisis periods.

We find that the coefficients of InvTV and IHP are significantly positive, indicating that order processing costs, inventory costs and adverse selection costs are major components of market makers' bid-ask spreads. This is consistent with the findings in BSW (2004) as well as our findings in Chapter 3. The coefficients of the interaction terms with non-crisis period dummies are found to be significantly negative, which suggests significant decreases in the coefficients on all cost components of the bid-ask spread during non-crisis periods relative to the crisis period. The results are consistent for all four models.

The coefficient of a cost component (InvTV and IHP) indicates the approximate amount of compensation a market maker would demand for one unit of such cost. While we witness significant decreases in the coefficients on InvTV and IHP during the non-crisis period, the coefficients become significantly higher during the GFC period. This indicates that market makers behave differently in setting their bid-ask spreads to compensate their costs with the impact of the GFC. Our findings suggest that market makers demand significantly higher compensation for the costs they are bearing during the GFC period, and demand relatively lower compensation during the non-crisis period.

Moreover, comparing the adjusted R-squares of models 1 and 3 with the ones of models 2 and 4, we find by adding InvTV to the regression models, the explanatory power merely increases by approximately 0.1%. This suggests that a market maker concentrates on inventory costs and adverse selection costs when setting his bid-ask spread, rather than on order processing costs.

[Insert Table 4.3]

We then estimate the PI in each sample period using the restricted BSW model. The model is described in section 4.2.4. The results are reported in Table 4.4. Panel A provides the results for the two regressions on EWQS, and Panel B provides the results for the two regressions on VWES. Models 1 and 3 contain variables including InvTV, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$, while models 2 and 4 only contain variables including $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. The regressions are run separately in each sample period.

According to section 4.2.4, PI is defined as the probability that a trade is made with an informed trader. Since trading with an informed trader causes adverse selection, a higher PI suggests a higher level of adverse selection in certain periods. In Table 4.4, we find significantly higher PIs during the crisis period and significantly lower PIs during the non-crisis periods. The results are consistent for all four models. The EWQS regressions estimate the PI during the crisis period to be approximately 7.9%, while the PIs during the pre-GFC and post-GFC periods are approximately 6% and 1.1%, respectively. The higher PI during the crisis period indicates higher adverse selection during the financial crisis. In addition, we find PI is significantly reduced after the crisis, and becomes a lot lower than the PI before the crisis, which indicates lower adverse selection when the market starts to recover. Similar results can be observed from the regressions on VWES. Moreover, we are still able to find evidence that InvTV does not add much explanatory power to the bid-ask spread in all four restricted models across the whole sample period, which is consistent with our findings on the modified BSW model.

Overall, our findings on the pool regression and the PI estimation reveal higher adverse selection on a cross-section of stocks during the crisis period.

4.4.3 Adverse Selection on Individual Stocks

We estimate daily PIs of individual stocks by solving Equation (4.11) and (4.12), in which IHP is the probability weighted average of $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. Table 4.5 presents the average daily PI of individual sample stocks in all three periods. Panel A reports the average daily PIs of individual stocks, which are estimated with equal-weighted IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. Panel B reports the average daily PIs of individual stocks, which are estimated with volume-weighted IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$.

According to Panel A, 21 of the 25 sample stocks experience a PI increase during the crisis period, while 18 experience a PI decrease after the crisis. Most of the sample stocks have higher PIs during the crisis period, which suggests higher adverse selection on a stock level. Similar results can be found on Panel B. Our findings suggest that most stocks experienced significant increases in PIs during the GFC period, indicating higher adverse selection during the crisis period on a stock level.

[Insert table 4.5]

We then use the estimated daily PIs of individual stocks as a dependent variable and run a regression on the GFC period dummy. Results are reported in Table 4.6. Panel A employs the daily PI estimated with the equal-weighted average of IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$, and Panel B employs the daily PI estimated with the volume-weighted average of IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$. Both panels show a significantly positive coefficient of the GFC period dummy, indicating a significant increase in the daily PI of individual stocks during the GFC period. Our findings suggest higher adverse selection during the crisis period at an individual stock level, which is consistent with our findings on a market level.

[Insert table 4.6]

4.5 Conclusion

This chapter examines the impact of the GFC on market makers' behaviours and the level of adverse selection in stock trading.

We found that market makers are more conservative in setting their bid-ask spreads during crisis periods. They demand significantly higher compensation for their costs in the GFC period than in the non-GFC period. As a result, the bid-ask spreads widen significantly during the crisis period. Moreover, comparing the R-squares of models with and without the order processing cost component reveals that market makers are less concerned with order processing costs, than with inventory and adverse selection costs.

We estimate PI for a cross-section of stocks and use it as an indicator of the level of adverse selection on a market level, during the crisis and non-crisis periods. We observed a significantly higher level of adverse selection during the crisis period relative to the non-crisis period, and a significant decrease in adverse selection when the market is recovering from the crisis. We also introduce a method to estimate daily PI for individual stocks with intraday observations. We used the proposed method to estimate daily PIs of individual sample stocks with during the crisis and non-crisis periods. We found that most stocks faced higher adverse selection during the crisis period, consistent with our findings at a market level.

Our proposed method of PI estimation allows us to identify PI of individual stocks on a single trading day as well as in a certain intraday period. The new PI estimation per stock per day allows us to monitor informed trading in certain stocks at the time of takeovers, earning announcements and other significant events. This offers great potential for future research and provides regulators and researchers with new and effective tools for dealing with informed trading.

Table 4.1 Summary statistics in pre-GFC, GFC and post-GFC periods

This table presents the summary of descriptive statistics of variables for sample stocks in intraday calendar clock intervals (10-minute) during the sample period. The sample period covers pre-GFC period (August 2005), GFC period (August 2007) and post-GFC period (August 2009). The sample contains 25 stocks, which are the stocks included in the Dow Jones 30 stocks during the whole sample period. EWQS is the equal-weighted quoted bid-ask spread; VWES is the volume-weighted effective bid-ask spread; REWQS is EWQS divided by stock price; RVWES is VWES divided by stock price; S is the average stock price in each interval; InvS is the inverse of S; TV is the number of shares traded in each interval; InvTV is the inverse of TV; NT is the number of transactions in each interval; σ is the annualized realized volatility of stock return in each interval; $\overline{\sqrt{t}}$ is the average of the square root of time between transactions; and IHP is the inventory holding premium.

Panel A. Pre-GFC Period (August 2005)					
Variable	nobs	mean	25%	median	75%
Spread measures					
EWQS	22424	0.0921	0.0354	0.0645	0.1178
VWES	22424	0.0379	0.0132	0.0218	0.0420
REWQS	22424	0.0020	0.0011	0.0016	0.0024
RVWES	22424	0.0009	0.0004	0.0005	0.0009
Determinants of spread					
S	22424	43.03	27.58	40.43	55.16
InvS	22424	0.0268	0.0181	0.0247	0.0363
TV	22424	135,958	57,200	101,700	174,800
InvTV	22424	1.3759E-05	5.721E-06	9.8328E-06	1.7483E-05
NT	22424	145	95	129	180
$\sqrt{t}$	22424	0.0008	0.0007	0.0008	0.0009
σ	22424	0.1726	0.1215	0.1574	0.2083
IHP	22424	0.0023	0.0013	0.0021	0.0031

Panel B. GFC Period (August 2007)					
Variable	nobs	mean	25%	median	75%
Spread measures					
EWQS	22424	0.1538	0.0442	0.0928	0.2104
VWES	22424	0.0966	0.0221	0.0473	0.1269
REWQS	22424	0.0026	0.0010	0.0019	0.0036
RVWES	22424	0.0017	0.0005	0.0009	0.0021
Determinants of spread					
S	22424	54.35	38.10	48.23	65.15
InvS	22424	0.0217	0.0153	0.0207	0.0262
TV	22424	82,930	38,950	60,300	97,000
InvTV	22424	1.9726E-05	1.031E-05	1.6584E-05	2.5674E-05
NT	22424	228	175	223	275
$\sqrt{t}$	22424	0.0006	0.0006	0.0006	0.0007
σ	22424	0.2750	0.2115	0.2583	0.3208
IHP	22424	0.0038	0.0024	0.0031	0.0048

Panel C. Post-GFC Period (August 2009)					
Variable	nobs	mean	25%	median	75%
Spread measures					
EWQS	20474	0.0249	0.0121	0.0165	0.0277
VWES	20474	0.0155	0.0092	0.0107	0.0160
REWQS	20474	0.0006	0.0004	0.0005	0.0007
RVWES	20474	0.0004	0.0002	0.0003	0.0005
Determinants of spread					
S	20474	42.21	26.06	42.30	53.48
InvS	20474	0.0314	0.0187	0.0236	0.0384
TV	20474	69,179	31,155	46,736	80,305
InvTV	20474	2.449E-05	1.245E-05	2.1397E-05	3.2098E-05
NT	20474	230	169	219	278
$\sqrt{t}$	20474	0.0006	0.0006	0.0006	0.0007
σ	20474	0.2704	0.1880	0.2367	0.3525
IHP	20474	0.0027	0.0015	0.0023	0.0034

Table 4.2 Summary of cross-correlations in pre-GFC, GFC and post-GFC periods

This table presents the cross-correlations between variables for sample stocks in intraday calendar clock intervals (10-minute) during the sample period. The sample period covers pre-GFC period (August 2005), GFC period (August 2007) and post-GFC period (August 2009). The sample contains 25 stocks, which are the stocks included in the Dow Jones 30 stocks during the whole sample period. EWQS is the equal-weighted quoted bid-ask spread; VWES is the volume-weighted effective bid-ask spread; REWQS is EWQS divided by stock price; RVWES is VWES divided by stock price; S is the average stock price in each interval; InvS is the inverse of S; TV is the number of shares traded in each interval; InvTV is the inverse of TV; NT is the number of transactions in each interval; σ is the annualized realized volatility of stock return in each interval; $\sqrt{t}$ is the average of the square root of time between transactions; IHP is the inventory holding premium; IHP_I is the inventory holding premium for transactions with informed traders.

Panel A												
VARIABLE	EWQS	VWES	REWQS	RVWES	S	InvS	TV	InvTV	NT	$\sqrt{t}$	σ	IHP
VWES	0.6474											
REWQS	0.8903	0.6104										
RVWES	0.4983	0.9160	0.5997									
S	0.5663	0.3157	0.2281	0.0608								
InvS	-0.5328	-0.3174	-0.2273	-0.0718	-0.9379							
TV	-0.2265	-0.0932	-0.1618	-0.0256	-0.2740	0.2288						
InvTV	0.2012	0.0747	0.1659	0.0394	0.2167	-0.1208	-0.5745					
NT	-0.2526	-0.1336	-0.2371	-0.0975	-0.2374	0.1838	0.7075	-0.6115				
$\sqrt{t}$	0.1679	0.0770	0.1760	0.0677	0.1358	-0.0349	-0.6115	0.7819	-0.8772			
σ	-0.1411	-0.0724	-0.0502	0.0006	-0.2658	0.2227	0.1220	-0.1024	0.1707	-0.1683		
IHP	0.4028	0.2224	0.2148	0.0725	0.6300	-0.6070	-0.3887	0.4240	-0.4259	0.3805	0.4146	
IHP_I	0.5681	0.3169	0.2301	0.0619	1.0000	-0.9379	-0.2742	0.2169	-0.2377	0.1360	-0.2658	0.6302

Panel B

VARIABLE	EWQS	VWES	REWQS	RVWES	S	InvS	TV	InvTV	NT	$\sqrt{t}$	σ	IHP
VWES	0.7712											
REWQS	0.8835	0.7629										
RVWES	0.6249	0.9343	0.7652									
S	0.5634	0.3348	0.2101	0.0770								
InvS	-0.5212	-0.3484	-0.2499	-0.1325	-0.8988							
TV	-0.2180	-0.1489	-0.0998	-0.0512	-0.3829	0.4838						
InvTV	0.2542	0.1705	0.0914	0.0425	0.5040	-0.5383	-0.5253					
NT	-0.1204	-0.1007	0.0079	-0.0057	-0.3153	0.3627	0.4970	-0.7627				
$\sqrt{t}$	0.0918	0.0796	-0.0311	-0.0095	0.3147	-0.3500	-0.4414	0.8554	-0.9387			
σ	-0.0534	-0.0142	0.0263	0.0243	-0.1538	0.0906	0.0282	-0.1140	0.1118	-0.1209		
IHP	0.4106	0.2725	0.1680	0.0845	0.7532	-0.7103	-0.3779	0.5618	-0.4335	0.4551	0.4245	
IHP_I	0.5657	0.3369	0.2128	0.0790	1.0000	-0.8988	-0.3829	0.5039	-0.3152	0.3145	-0.1538	0.7531

Panel C

VARIABLE	EWQS	VWES	REWQS	RVWES	S	InvS	TV	InvTV	NT	$\sqrt{t}$	σ	IHP
VWES	0.8544											
REWQS	0.7130	0.6162										
RVWES	0.4033	0.6608	0.6997									
S	0.5378	0.4015	-0.0821	-0.2228								
InvS	-0.3974	-0.2823	0.1656	0.3236	-0.8044							
TV	-0.2118	-0.1357	0.0992	0.2161	-0.4227	0.6003						
InvTV	0.2870	0.1766	-0.0200	-0.1587	0.4873	-0.5191	-0.5921					
NT	-0.1729	-0.0935	0.0722	0.1740	-0.3369	0.4214	0.6251	-0.7742				
$\sqrt{t}$	0.1734	0.0898	-0.0615	-0.1592	0.3357	-0.3788	-0.5356	0.8800	-0.9300			
σ	-0.0452	-0.0268	0.3157	0.2925	-0.4027	0.4808	0.2475	-0.1465	0.1524	-0.1106		
IHP	0.5040	0.3667	0.1137	-0.0686	0.6834	-0.5744	-0.4124	0.6428	-0.5012	0.5581	0.2239	
IHP_I	0.5383	0.4020	-0.0817	-0.2224	1.0000	-0.8044	-0.4226	0.4873	-0.3369	0.3357	-0.4026	0.6834

Table 4.3 Regression results on market makers' behaviours

This table presents the results of the cross-sectional regression of the bid-ask spreads of sample stocks using the modified BSW model. EWQS is the equal-weighted quoted bid-ask spread; VWES is the volume-weighted effective bid-ask spread; InvTV is the inverse of number of shares traded; and IHP is the inventory holding premium. d_1 and d_2 are dummy variables for pre-GFC period and post-GFC period respectively. d_1 is 1 for pre-GFC period, and 0 otherwise. d_2 is 1 for post-GFC period, and 0 otherwise. The sample period covers pre-GFC period (August 2005), GFC period (August 2007) and post-GFC period (August 2009).

Panel A contains the results of 2 regressions of EWQS:

Model 1: $\mathrm{EWQS} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_2 \mathrm{InvTV} \times \mathrm{d}_1 + \alpha_3 \mathrm{InvTV} \times \mathrm{d}_2 + \alpha_4 \mathrm{IHP} + \alpha_5 \mathrm{IHP} \times \mathrm{d}_1 + \alpha_6 \mathrm{IHP} \times \mathrm{d}_2 + \varepsilon$

Model 2: $\mathrm{EWQS} = \alpha_0 + \alpha_4 \mathrm{IHP} + \alpha_5 \mathrm{IHP} \times \mathrm{d}_1 + \alpha_6 \mathrm{IHP} \times \mathrm{d}_2 + \varepsilon$

Panel B contains the results of 2 regressions of VWES:

Model 3: $\mathrm{VWES} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_2 \mathrm{InvTV} \times \mathrm{d}_1 + \alpha_3 \mathrm{InvTV} \times \mathrm{d}_2 + \alpha_4 \mathrm{IHP} + \alpha_5 \mathrm{IHP} \times \mathrm{d}_1 + \alpha_6 \mathrm{IHP} \times \mathrm{d}_2 + \varepsilon$

Model 4: $\mathrm{VWES} = \alpha_0 + \alpha_4 \mathrm{IHP} + \alpha_5 \mathrm{IHP} \times \mathrm{d}_1 + \alpha_6 \mathrm{IHP} \times \mathrm{d}_2 + \varepsilon$

Panel A									
		Coefficient estimates / t-stat							
EWQS	Nobs	Intercept	InvTV	$\mathrm{InvTV} \times \mathrm{d}_1$	$\mathrm{InvTV} \times \mathrm{d}_2$	IHP	$\mathrm{IHP} \times \mathrm{d}_1$	$\mathrm{IHP} \times \mathrm{d}_2$	R^2
Model 1	65322	0.0179	515.5495	-249.8163	-757.2382	32.6096	-2.3588	-26.7285	0.3315
		22.0545	9.0502	-3.1386	-9.9182	95.2280	-4.4727	-47.1544	
Model 2	65322	0.0184				35.0425	-3.4548	-31.2776	0.3302
		23.2403				157.2566	-11.1513	-119.4587	

Panel B									
		Coefficient estimates / t-stat							
VWES	Nobs	Intercept	InvTV	$\mathrm{InvTV} \times \mathrm{d}_1$	$\mathrm{InvTV} \times \mathrm{d}_2$	IHP	$\mathrm{IHP} \times \mathrm{d}_1$	$\mathrm{IHP} \times \mathrm{d}_2$	R^2
Model 3	65322	0.0166	356.0576	-458.6124	-572.4745	18.6465	-8.7893	-16.1913	0.2100
		25.2117	7.6873	-7.0865	-9.2220	66.9702	-20.4975	-35.1314	
Model 4	65322	0.0166				20.4064	-11.1155	-19.7173	0.2089
		25.7945				112.6589	-44.1379	-92.6439	

Table 4.4 Estimation of probability of informed trading on a market level

This table presents results for the PI estimation from the restricted regression with a cross-section of sample stocks. The sample contains 25 stocks, which are the stocks included in the Dow Jones 30 stocks during the whole sample period. The sample period covers pre-GFC period (August 2005), GFC period (August 2007) and post-GFC period (August 2009). The regressions are run in each period separately. Equal-weighted quoted spread (EWQS) and volume-weighted effective spread are employed as the spread measure; InvTV is the inverse of number of shares traded; $\mathrm{IHP_I}$ is inventory holding premium for transactions with informed traders; $\mathrm{IHP_U}$ is the inventory holding premium for transaction with uninformed traders; and PI is the probability of informed during each period. Panel A employs EWQS as the spread measure, and Panel B employs VWES as the spread measure.

Panel A contains the results of 2 regressions of EWQS:

Model 1: $\mathrm{EWQS} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_2 \mathrm{IHP_U} + \alpha_2 \times \mathrm{PI} \times (\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon$

Model 2: $\mathrm{EWQS} = \alpha_0 + \alpha_2 \mathrm{IHP_U} + \alpha_2 \times \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon$

Panel B contains the results of 2 regressions of VWES:

Model 3: $\mathrm{VWES} = \alpha_0 + \alpha_1 \mathrm{InvTV} + \alpha_2 \mathrm{IHP_U} + \alpha_2 \times \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon$

Model 4: $\mathrm{VWES} = \alpha_0 + \alpha_2 \mathrm{IHP_U} + \alpha_2 \times \mathrm{PI}(\mathrm{IHP_I} - \mathrm{IHP_U}) + \varepsilon$

with the restriction of $\alpha_2 = 1$.

Panel A							
			Coefficient estimates / t-stat				
EWQS	Year	Nobs	Intercept	InvTV	$\mathrm{IHP_I} - \mathrm{IHP_U}$	$\mathrm{IHP_U}$	R^2
Model 1	2005	22424	-0.0420	575.3834	0.0586	1.00	0.3289
			-30.6618	14.5886	98.2169		
	2007	22424	-0.0526	-478.3948	0.0793	1.00	0.3207
			-23.7904	-6.5030	92.0393		
	2009	20474	-0.0003	37.6150	0.0114	1.00	0.2836
			-1.0329	3.6649	77.1884		
Model 2	2005	22424	-0.0381		0.0605	1.00	0.3225
			-28.2487		103.3666		
	2007	22424	-0.0544		0.0765	1.00	0.3194
			-24.7437		102.6709		
	2009	20474	0.0000		0.0117	1.00	0.2831
			0.0956		90.4046		

Panel B							
			Coefficient estimates / t-stat				
VWES		Nobs	Intercept	InvTV	$\mathrm{IHP_I - IHP_U}$	$\mathrm{IHP_U}$	R^2
Model 3	2005	22424	-0.0041	25.1207	0.0193	1.00	0.1003
			-4.4699	0.9577	48.6613		
	2007	22424	-0.0058	8.6702	0.0375	1.00	0.1129
			-2.7650	0.1250	46.1814		
	2009	20474	0.0036	-36.4527	0.0059	1.00	0.1533
			15.5807	-4.9754	56.3110		
Model 4	2005	22424	-0.0039		0.0194	1.00	0.1002
			-4.3672		50.0610		
	2007	22424	-0.0057		0.0376	1.00	0.1129
			-2.7705		53.5411		
	2009	20474	0.0033		0.0057	1.00	0.1523
			14.7578		61.6696		

Table 4.5 Estimation of probability of informed trading on a stock level

This table presents results for average daily probability of informed trading (PI) estimation for individual sample stocks. For a single trading day, PI of an individual stock is inferred by the following model:

$$\mathrm{IHP} = \mathrm{PI} \times \mathrm{IHP_I} + (1 - \mathrm{PI}) \times \mathrm{IHP_U}$$

The sample period covers pre-GFC period (August 2005), GFC period (August 2007) and post-GFC period (August 2009). The sample contains 25 stocks, which are the stocks included in the Dow Jones 30 stocks during the whole sample period.

Panel A contains the results of daily PI estimation from model 1:

Model 1: $\mathrm{EWIHP} = \mathrm{PI} \times \mathrm{EWIHP_I} + (1 - \mathrm{PI}) \times \mathrm{EWIHP_U}$

where EWIHP, $\mathrm{EWIHP_I}$ and $\mathrm{EWIHP_U}$are equal-weighted average of IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$ respectively, and PI is the daily probability of informed trading.

Panel B contains the results of daily PI estimation from model 2:

Model 2: $\mathrm{VWIHP} = \mathrm{PI} \times \mathrm{VWIHP_I} + (1 - \mathrm{PI}) \times \mathrm{VWIHP_U}$

where VWIHP, $\mathrm{VWIHP_I}$ and $\mathrm{VWIHP_U}$are volume-weighted average of IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$ respectively, and PI is the daily probability of informed trading.

Panel A							
	Probability of informed trading				Probability of informed trading		
SYMBOL	2005	2007	2009	SYMBOL	2005	2007	2009
AA	0.1636%	0.2268%	0.2238%	KO	0.0887%	0.0989%	0.0758%
AXP	0.0905%	0.2426%	0.1935%	MCD	0.1851%	0.1095%	0.0928%
BA	0.1120%	0.1532%	0.1542%	MMM	0.0846%	0.0886%	0.1537%
CAT	0.1740%	0.1624%	0.1974%	MRK	0.1344%	0.1659%	0.1252%
DD	0.1700%	0.1758%	0.1682%	MSFT	0.0956%	0.1056%	0.1291%
DIS	0.1218%	0.1320%	0.1628%	PFE	0.0780%	0.0936%	0.0976%
GE	0.0590%	0.1174%	0.1592%	PG	0.0708%	0.0962%	0.0909%
HD	0.1168%	0.1567%	0.1075%	T	0.1162%	0.1256%	0.0851%
HPQ	0.1667%	0.1407%	0.0923%	UTX	0.1168%	0.1352%	0.1344%
IBM	0.0936%	0.1235%	0.1018%	VZ	0.0702%	0.1153%	0.1007%
INTC	0.0985%	0.1311%	0.1022%	WMT	0.0692%	0.1329%	0.0574%
JNJ	0.1031%	0.0824%	0.0674%	XOM	0.1003%	0.1522%	0.0810%
JPM	0.0643%	0.1548%	0.1228%				

Panel B							
	Probability of informed trading				Probability of informed trading		
SYMBOL	2005	2007	2009	SYMBOL	2005	2007	2009
AA	0.1455%	0.1639%	0.2067%	KO	0.0846%	0.0784%	0.0634%
AXP	0.0815%	0.1587%	0.1506%	MCD	0.1537%	0.0920%	0.0694%
BA	0.0951%	0.1145%	0.1213%	MMM	0.0738%	0.0758%	0.1152%
CAT	0.1386%	0.1278%	0.1538%	MRK	0.1200%	0.1211%	0.1068%
DD	0.1361%	0.1286%	0.1347%	MSFT	0.0861%	0.0940%	0.1182%
DIS	0.1213%	0.1065%	0.1360%	PFE	0.0772%	0.0902%	0.0939%
GE	0.0589%	0.0937%	0.1535%	PG	0.0697%	0.0735%	0.0758%
HD	0.1081%	0.1195%	0.0946%	T	0.1035%	0.1004%	0.0795%
HPQ	0.1502%	0.1025%	0.0780%	UTX	0.1039%	0.1039%	0.1059%
IBM	0.0782%	0.0951%	0.0777%	VZ	0.0696%	0.0944%	0.0891%
INTC	0.0904%	0.1152%	0.0965%	WMT	0.0690%	0.1009%	0.0510%
JNJ	0.0919%	0.0647%	0.0573%	XOM	0.0883%	0.1027%	0.0638%
JPM	0.0646%	0.1104%	0.1006%				

Table 4.6 Regression results on adverse selection of individual stocks in the crisis

This table presents the results of the regression of the daily probability of informed trading (PI) of sample stocks on the GFC-period dummy. PI is the estimated daily probability of informed trading for an individual stock and $\mathrm{d_{GFC}}$ is the dummy variable for the GFC period. $\mathrm{d_{GFC}}$ is 1 for GFC period, and 0 otherwise. The sample period covers pre-GFC period (August 2005), GFC period (August 2007) and post-GFC period (August 2009). The sample contains 25 stocks, which are the stocks included in the Dow Jones 30 stocks during the whole sample period.

Panel A contains the results of model 1:

Model 1: $\mathrm{PI} = \alpha_0 + \alpha_1 \mathrm{d_{GFC}} + \varepsilon$

where PI is estimated with equal-weighted average of IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$.

Panel B contains the results of model 1:

Model 1: $\mathrm{PI} = \alpha_0 + \alpha_1 \mathrm{d_{GFC}} + \varepsilon$

where PI is estimated with volume-weighted average of IHP, $\mathrm{IHP_I}$ and $\mathrm{IHP_U}$.

Panel A				
		Coefficient estimates / t-stat		
PI	Nobs	Intercept	$\mathrm{d_{GFC}}$	R^2
Model 1	1675	0.00116	0.00021	0.04869
		88.85808	9.25338	

Panel B				
		Coefficient estimates / t-stat		
PI	Nobs	Intercept	$\mathrm{d_{GFC}}$	R^2
Model 1	1675	0.00101	0.00004	0.00380
		103.62133	2.52672	

Chapter 5

CONCLUSION

The thesis starts by introducing a different way to measure time in finance studies using event clocks. We construct an event clock, termed “transaction clock”, in which the event is defined as the occurrence of a constant number of transactions. The transaction clock is found to have good characteristics, such as capturing the same level of market activity and uncovering a normal distribution of intraday stock returns. We then apply the transaction clock to examine the intraday bid-ask spread cost components in a high-frequency market. We employ the BSW model, which assumes a normal distribution of stock returns. Our findings suggest the BSW model explains the bid-ask spread in a high-frequency market. In addition, the explanatory power of the model is improved by applying the transaction clock, as it fulfils the model’s underlying assumption. Based on the modified BSW model, we develop one method of PI estimation on a market level using a restricted regression with intraday observations, and another method of PI estimation of individual stocks on a daily basis by solving a fully identified equation. We apply both methods of PI estimation to examine the impact of the GFC on market makers’ behaviours and adverse selection in stock trading.

In the first essay (Chapter 2), we introduce a different way to measure time changes with a transaction clock and a volume clock. The transaction clock and the volume clock are found to perform better than the traditional calendar clock in capturing the level of market activity. The intraday stock returns are found to follow a normal distribution when we use a transaction clock to measure time changes. On the other hand, intraday returns do not follow a normal distribution within the setting of the traditional calendar clock, but follow a mixed distribution of a zero-mean normal distribution and a scaled, non-negative and positively skewed distribution of number of transactions (or volume).

Our findings in the first essay suggest that the transaction clock possesses good characteristics for capturing a constant level of market activity and uncovering a normal distribution of stock returns. The transaction clock brings new insight and great potential for finance studies, despite the fact that most extant literature uses traditional time measurement with a calendar clock. Few finance theories actually specify a particular time measurement; nevertheless, the consistent characteristics of the transaction clock make it especially suited to finance studies. Several financial models assume a normal distribution of stock returns; thus, empirical tests with a transaction clock are likely to provide more meaningful results by fulfilling the underlying assumption.

The second essay (Chapter 3) employs the BSW model to re-evaluate the intraday cost components of market makers' bid-ask spreads in a high-frequency market. The model assumes a normal distribution of stock returns. We therefore conduct the empirical tests with a transaction clock verses a calendar clock, in order to examine the benefits of applying a transaction clock in this study.

Our findings suggest that market makers behave rationally under current market conditions, despite more frequent posting of bid and ask quotations and delivering of transactions. Accordingly, the BSW model possesses good explanatory power in the current high-frequency market. Specifically, inventory costs and adverse selection costs have significant impact on market makers' bid-ask spreads, while the impact of order processing costs is not as significant. In a high-frequency market, market makers are compensated for their inventory and adverse selection costs by a significant portion of their bid-ask spreads. Moreover, quoted spreads perform better than effective spreads in improving the explanatory power of the model.

Comparing the empirical results within the setting of a transaction clock with the results within the setting of a calendar clock reveals that applying a transaction clock significantly improves the explanatory power of the model. While the coefficients on the independent variables are consistent across all the models with both clocks, the transaction clock outperforms the traditional calendar clock in achieving significantly higher R-squares. Our findings in the first and second essays suggest there is great benefit in applying the transaction clock in financial studies, especially when a model assumes normality of returns distribution.

Based on the improved model, we introduce a method to estimate PI on a market level using a restricted regression with intraday observations. Our proposed method has advantages over the method in the original study in identifying PI over certain short periods, such as an hour, a day and a week. Our method allows us to more accurately identify the impact of asymmetric information and root out insider trading activity.

In the third essay (Chapter 4), we apply the PI estimation method on a market level to examine the impact of the GFC on market makers' behaviours and the level of adverse selection in the stock market. In addition, we develop a new method of PI estimation on a stock level by solving a fully identified equation, and apply the method to examine the adverse selection of individual stocks during the crisis and non-crisis periods.

We find market makers are more conservative in setting their bid-ask spread during the crisis period, than during non-crisis periods. They tend to widen their bid-ask spreads and demand significantly higher compensation for their costs in the GFC period. Moreover, we find that market makers are less concerned about order processing costs, than inventory and adverse selections costs, which is consistent with the findings in the second essay.

We estimate PI for a cross-section of stocks during the crisis and non-crisis periods using a restricted regression. By comparing the PIs in different periods, we find a significant increase in adverse selection on a market level during the crisis period, followed by a significant drop in adverse selection when the market recovers. We then estimate daily PI for individual stocks during the sample period by solving a fully identified equation. By examining daily PIs of individual stocks, we find that the arrival of the GFC significantly increases adverse selection for individual stocks, which is consistent with our findings on a market level.

Our proposed methods of PI estimation on a market level and a stock level provide important new tools for regulators, researchers, investors and other parties to understand market behaviour and deal with informed trading. The PI estimation on a market level identifies the level of adverse selection in the stock market in certain periods. It allows us to understand market behaviour and investor confidence around key market-related events, such as wars and crisis. The PI estimation on a stock level allows us to identify PI of individual stocks on a single trading day as well as in a certain intraday period. With this method, we can monitor informed trading in certain stocks during takeovers, earning announcements and other significant events, offering great potential for future research.

LIST OF REFERENCES

Affleck-Graves, J., and W. McDonald, 1989. Non-normalities and tests of asset pricing theories, *Journal of Finance* 44, 889-908.

Andersen T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001. The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43-76.

Admati, A.R., 1985. A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica* 53, 629-658.

Admati, A.R. & Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. *The Review of Financial Studies* 1, 3-40.

Ané, T., and H. Geman, 2000. Order flow, transaction clock and normality of asset returns, Journal of Finance 55, 2259-2285.

Bachelier, L., 1900. Théorie de la speculation, *Annales Scientifiques de 1'Ecole Normale Supérueyre* 17, 21-86.

Black, F., and M. Scholes, 1973. The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637-659.

Brailsford, T.J., 1996. The empirical relationship between trading volume, returns and volatility, *Accounting and Finance* 36, 89-111.

Branch, B. & Freed, W., 1977. Bid-ask spreads on AMEX and the big board. *Journal of Finance* 32, 159-163.

Bollen, N.P.B., Smith, T. & Whaley, R.E., 2004. Modeling the bid/ask spread: measuring the inventory-holding premium. *Journal of Financial Economics* 72, 97-141.

Boulatov, A. and George, T. J., 2013. Hidden and displayed liquidity in securities markets with informed liquidity providers. *The Review of Financial Studies* 26, 2096-2137.

Chan, K.C., Christie, W.G., and Schultz, P.H., 1995. Market structure and the intraday pattern of bid-ask spreads for NASDAQ securities. *The Journal of Business* 68, 35-60.

Cheung, Y., R.Y. Ho, P. Pope, and P. Draper, 1994. Intraday stock return volatility: the Hong Kong evidence, *Pacific-Basin Finance Journal* 2, 261-276.

Choe, H., and H.S. Shin, 1993. An analysis of interday and intraday return volatility – evidence from the Korea Stock Exchange, *Pacific-Basin Finance Journal* 1, 175-188.

Clark, P.K., 1973. A subordinated stochastic process with finite variance of speculative prices, *Econometrica* 41, 135-156.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985. An intertemporal general equilibrium model of asset prices, *Econometrica* 53, 363-384.

Doan, M.P., C.T. Lin, and M. Chng, 2014. Higher moments and beta asymmetry: evidence from Australia, *Accounting and Finance* 54, 779-807.

Easley, D., M. Lopez de Prado, and M. O'Hara, 2012. Flow toxicity and liquidity in a high-frequency world, *The Review of Financial Study* 25, 1457-1493.

Easley, D., Kiefer, N.M., O'Hara, M. & Paperman, J.B., 1996. Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51, 1405-1436.

Easley, D. and O'Hara, M., 1987. Price, trade size, and information in securities markets. *Journal of Financial Economics* 19, 69-90.

Fama, E.F., 1965. The behavior of stock market prices, *Journal of Business* 38, 34-105.

Foster, F.D. & Viswanathan, S., 1990. A theory of the interday variation in volume, variance, and trading costs in securities markets. *The Review of Financial Studies* 1990, 593-624.

Geman, H., 2005. From measure changes to time changes in asset pricing, *Journal of Banking & Finance* 29, 2701-2722.

Glosten, L.R. & Harris, L., 1988. Estimating the components of the bid-ask spread. *Journal of Financial Economics* 21, 123-142.

Gloston, L.R. & Milgrom P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71-100.

Gray, P., E. Kalotay, and J. Mclvor, 1998. Testing multivariate normality of Australian stock returns, *Australian Journal of Management*, 23, 135-150

Grossman, S., 1976. On the efficiency of competitive stock markets where trades have diverse information. *Journal of Finance* 31, 573-585.

Grossman, S. & Stiglitz, J., 1980. On the impossibility of informationally efficient markets. *American Economic Review* 48, 541-559.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-1054.

Harris, L., 1994. Minimum price variations, discrete bid–ask spreads, and quotation sizes. *The Review of Financial Studies* 7, 149–178.

Hodgson, A., 1996. Information transfer, microstructures and intraday price return spikes, *Accounting and Finance* 36, 229-257.

Ke, M.C., C.H. Jiang, and Y.S. Huang, 2004. The impact of tick size on intraday stock price behavior: evidence from the Taiwan Stock Exchange, *Pacific-Basin Finance Journal* 12, 19-39.

Kwan, A., Masulis, R. and McInish, T.H., 2015. Trading rules, competition for order flow and market fragmentation. *Journal of Financial Economics* 115, 330-348.

Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315-1336.

Li, Y. Van Ness, B.F. and Van Ness R.A., 2005. Daily and intraday patterns in spread and depth for limit orders and specialists. *Quarterly Journal of Business and Economics* 44, 3-14.

Ling, X., 2015. Normality of stock returns with event time clocks. *Accounting and Finance*.

Lo, A.W., and A.G. MacKinlay, 1988. Stock market prices do not follow random walks: evidence from a simple specification test, *The Review of Financial Studies* 1, 41-66.

Mandelbrot, B., 1963. The variation of certain speculative prices, *Journal of Business* 36, 394-413.

Mandelbrot, B., and H. Taylor, 1967. On the distribution of stock prices differences, *Operations Research* 15, 1057-1062.

McInish, T.H. and Van Ness, B.F., 2002. An intraday examination of the components of the bid-ask spread. *The Financial Review* 37, 507-524.

McInish, T.H. and Wood R.A., 1992. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *Journal of Finance* 47, 753-764.

Merton, R.C., 1973. Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373-413.

Morris, S. and Shin, H.S., 2012. Contagious adverse selection. *American Economic Journal: Macroeconomics* 4(1), 1-21.

O'Hara, M., 2015. High frequency market microstructure. *Journal of Financial Economics* 116, 257-270.

Praetz, P., and E.J.G Wilson, 1978. The distribution of stock market returns: 1958-1973, *Australian Journal of Management* 3, 79-90.

Rosu, I., 2015. Liquidity and information in order driven markets. Working paper. http://ssrn.com/abstract=1286193.

Sidhu, B., Smith, T., Whaley, R.E., and Willis R.H., 2008. Regulation Fair Disclosure and the cost of adverse selection. Journal of Accounting Research 46(3), 697-728.

Silverman, B.W., 1986. *Density estimation for statistics and data analysis* (Chapman and Hall, London).

Stokie, M.D., 1982. The distribution of stock market returns: test of normality, *Australian Journal of Management* 7, 159-178.

Stoll, H.R., 1978. The pricing of security dealer services: an empirical study of Nasdaq stocks. *Journal of Finance*, 1153–1172.

Tannous G., Wang, J. and Wilson C., 2013. The intraday pattern of information asymmetry, spread, and depth: evidence from the NYSE. *International Review of Finance* 13, 215-240.

Tauchen, G.E., and M. Pitts, 1983. The price variability-volume relationship on speculative markets, *Econometrica* 51, 485-505.

Tinic, S., 1972. The economics of liquidity services. *Quarterly Journal of Economics* 86, 79-83.

Tinic, S., & West, R., 1972. Competition and the pricing of dealer services in the over-the-counter market. *Journal of Financial and Quantitative Analysis* 8, 1707-1727.

Upton, D.E., and D.S. Shannon, 1979. The stable Paretian distribution, subordinated stochastic processes and asymptotic log-normality: an empirical investigation, *Journal of Finance* 34, 1031-1039.

Zhu, Y. and Gippel, J., 2015. Informed trading on debt covenant violations: the long and short of it. *Accounting and Finance*, forthcoming.

APPENDICES

Appendix A

List of the sample Stocks in Chapter 2 and Chapter 3:

Ticker	Company
AXP	American Express Co
BA	Boeing Co
CAT	Catepillar Inc
CSCO	Cisco Systems Inc
CVX	Chevron Corp
DD	E I du Pont de Nemours and Co
DIS	Walt Disney Co
GE	General Electric Co
GS	Goldman Sachs Group Inc
HD	Home Depot Inc
IBM	International Business Machines Corp
INTC	Intel Corp
JNJ	Johnson & Johnson
JPM	JPMorgan Chase and Co
KO	The Coca-Cola Co
MCD	McDonald's Corp
MMM	3M Co
MRK	Merk & Co Inc
MSFT	Microsoft Corp
NKE	Nike Inc
PFE	Pfizer Inc
PG	Procter & Gamble Co
T	AT&T Inc
TRV	Travelers Companies Inc
UNH	UnitedHealth Group Inc
UTX	United Technologies Corp
V	Visa Inc
VZ	Verizon Communications Inc
WMT	Wal-Mart Stores Inc
XOM	Exxon Mobil Corp

Appendix B

List of the sample stocks in Chapter 4:

Ticker	Company
AA	Alcoa Inc
AXP	American Express Co
BA	Boeing Co
CAT	Catepillar Inc
DD	E I du Pont de Nemours and Co
DIS	Walt Disney Co
GE	General Electric Co
HD	Home Depot Inc
HPQ	Hewlett-Packard Co
IBM	International Business Machines Corp
INTC	Intel Corp
JNJ	Johnson & Johnson
JPM	JPMorgan Chase and Co
KO	The Coca-Cola Co
MCD	McDonald's Corp
MMM	3M Co
MRK	Merk & Co Inc
MSFT	Microsoft Corp
PFE	Pfizer Inc
PG	Procter & Gamble Co
T	AT&T Inc
UTX	United Technologies Corp
VZ	Verizon Communications Inc
WMT	Wal-Mart Stores Inc
XOM	Exxon Mobil Corp